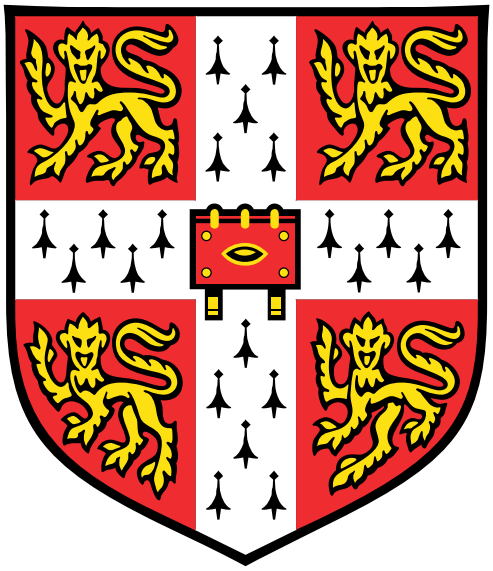
Pheromones in *Heliconius* butterflies:

Chemical ecology, genetics, and behaviour

Kathleen Darragh



Pembroke College, University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy in Zoology

September 2019

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This thesis does not exceed the prescribed word limit set by the Degree Committee for the School of Biology.

Acknowledgements

I am very lucky to have had an incredible mentor throughout my PhD. Chris Jiggins supported and encouraged me both when things were going well, but perhaps more importantly when times were difficult. My PhD has been truly enjoyable, full of special memories both in the field and in Cambridge. I thank Chris for his guidance, but also for the freedom he has given me to find my own direction as a scientist.

I acknowledge the funding I have received to support this work. My project was primarily funded by a studentship from Natural Environment Research Council, a short-term fellowship from the Smithsonian Tropical Research Institute (STRI), and fieldwork funding from the School of Biological Sciences. I thank all the academic and administrative staff at the Department of Zoology and STRI, as well as the staff and students of Pembroke College.

Being part of the Butterfly Genetics Group in Cambridge has been a wonderful experience. Whilst the people have changed over the years, the atmosphere has been consistently fun and supportive. I am especially grateful to Richard Merrill, Ian Warren and Kelsey Byers, who have always been willing to lend an ear and offer me their advice. I am very thankful for the other PhD students in the lab throughout my time: Gabriela Montejo-Kovacevich, Joe Hanly, Ana Pinharanda, Anna Orteu, and Henry North. Not only wonderful colleagues, and collaborators, but also friends who have made this experience infinitely more enjoyable.

I spent a significant amount of time in Panama at the STRI insectaries in Gamboa. This would not have been possible without Owen McMillan. I would like to thank Owen for his mentorship, he has been an immensely positive influence during this time. The insectary team became my family and support network while I was in Panama. In particular, Rachel Crisp, Sylvia Garza, Diana Abondano, and Oscar Paneso, who helped me not only with their hard work and dedication, but also whose company kept me sane throughout my field seasons. I am very grateful for all the wonderful people I met in Gamboa, it truly is a magical place and I will treasure the friendships that were formed there.

I thank my friends, both those from Garron Tower in Northern Ireland, and from my undergraduate at Pembroke, who helped me believe that I was capable of achieving this. I am also very grateful for my graduate friends from Pembroke, I'm not sure I would have made it without our Wednesday night tea and biscuits.

I thank my family for their never-ending and unconditional support. I would not be here today without the adventures which inspired me to become a biologist. Since you showed me how amazing nature could be I dreamed of going to the tropics, a dream that has now been fulfilled, and for that I am forever grateful.

Summary

Pheromones in Heliconius *butterflies: Chemical ecology, genetics, and behaviour*

Kathleen Darragh

Sex pheromones are chemical signals that mediate communication between males and females of a species and are important for mating. They have been well-studied in female moths, where females release volatile compounds to attract males over long distances. Butterflies fly during the day and were thought to rely more on vision. However, male pheromones in butterflies have also been described but have rarely been studied to the same detail as the female pheromones of night-flying moths.

The Neotropical butterflies, *Heliconius,* have a long history of ecological and genetic studies. Mate choice trials have focused on male preference for female colour pattern. In addition, two types of pheromone have been described in *Heliconius*. The first is produced by a region of the wing known as the androconia, and the second is anti-aphrodisiac, transferred from male genitals to females during mating to prevent re-mating by the female.

In this thesis I investigate the role of chemical signalling in mate choice of *Heliconius*. I describe sexual dimorphism in wing scale morphology and androconial chemical profile of *H. melpomene*. I demonstrate the importance of male chemical signalling for female choice in this and other *Heliconius* species. Male pheromones in other groups can convey information about mate quality. I show that larval but not adult diet affects the production of only minor components of the male pheromone. This suggests that male signalling cannot convey information about adult diet quality but could reflect larval diet. I survey variation in pheromone components across the broad geographic range of several species. I show that this provides a useful approach to identify biologically active components, which we expect to be species-specific over a large geographic range. There is also considerable interest in understanding the genetic basis for adaptive traits, and pheromones provide a tractable system for such studies. I

uncover the genetic basis for anti-aphrodisiac production in *H. melpomene* through genetic mapping and use in-vitro assays for enzymatic activity to identify a novel terpene synthase activity in a gene family not previously known for this role. This finding suggests that terpene synthesis has evolved independently multiple times in insects.

The work presented in this thesis is one of the first studies integrating both chemical ecology and genetics in butterflies. By combining behavioural, genetic and chemical studies, we can understand not only the composition of the chemical profile of individuals and species, but also their function and evolution.

Collaborations and publications

The four data chapters of this thesis have been written for publication, as individual manuscripts. This work has involved collaboration with others.

The majority of my fieldwork was carried out under the supervision of Owen McMillan at the Smithsonian Tropical Research Institute (STRI). STRI also funded some of the work through a Short-Term Fellowship. Another important collaboration was with Stefan Schulz at the Technische Universität Braunschweig, Institute of Organic Chemistry. Stefan aided with the chemical analysis, helping to develop the library of compounds used for sample analysis throughout this thesis.

In Chapter 2, I collected the behavioural and chemical data for *H. melpomene rosina*, conducted the analyses, and wrote the manuscript. Sohini Vanjari collected the chemical samples for an Ecuadorian population of *H. melpomene*, as well as taking the SEM images of the scales. Florian Mann and Stefan Schulz carried out chemical analysis. Maria Gonzalez-Rojas conducted behavioural trials in *H. timareta* and *H. melpomene malleti*, supervised by Camilo Salazar and Carolina Pardo-Diaz. Colin Morrison conducted behavioural trials in *H. erato*. Richard Merrill, Owen McMillan and Chris Jiggins contributed to the chapter concept and experimental design. All authors gave feedback on the manuscript.

Chapter 2 is published:

Darragh, K., S. Vanjari, F. Mann, M. F. Gonzalez-Rojas, C. R. Morrison, C. Salazar, C. Pardo-Diaz, R. M. Merrill, W. O. McMillan, S. Schulz, and C. D. Jiggins. 2017. Male sex pheromone components in *Heliconius* butterflies released by the androconia affect female choice. PeerJ 5:e3953.

In Chapter 3, I designed the experiment, collected the data, analysed the data, and wrote the manuscript, with help from Kelsey Byers, Richard Merrill, Owen McMillan and Chris Jiggins. Stefan Schulz assisted with the chemical analysis. All authors gave feedback on the manuscript.

Chapter 3 is published:

Darragh, K., K. J. R. P. Byers, R. M. Merrill, W. O. McMillan, S. Schulz, and C. D. Jiggins. 2019. Male pheromone composition depends on larval but not adult diet in *Heliconius melpomene*. Ecol Entomol, 44:397-405. doi: 10.1111/een.12716.

In Chapter 4, Gabriela Montejo-Kovacevich, Krzysztof Kozak, Colin Morrison, Owen McMillan, Chris Jiggins and I collected the samples. Clarisse Figueiredo, Jonathan Ready, Camilo Salazar and Mauricio Linares provided permits and assisted with trip organisation in Brazil and Colombia. I analysed the data and wrote the manuscript. Stefan Schulz assisted with the chemical analysis. Gabriela Montejo-Kovacevich created the intraspecific genetic distance matrices. Richard Merrill, Owen McMillan, Kelsey Byers and Chris Jiggins contributed to the chapter concept. All authors gave feedback on the manuscript.

Chapter 4 is available as a pre-print:

Darragh, K., G. Montejo-Kovacevich, K. M. Kozak, C. R. Morrison, C. M. E. Figueiredo, J. S. Ready, C. Salazar, M. Linares, K. J. R. P. Byers, R. M. Merrill, W. O. McMillan, S. Schulz, and C. D. Jiggins. 2019. Species specificity and intraspecific variation in the chemical profiles of *Heliconius* butterflies across a large geographic range. bioRxiv 573469. doi: 10.1101/573469

In Chapter 5, I collected the QTL data, with help from Sylvia Garza, Diana Abondano, and Kelsey Byers. Pasi Rastas created the linkage maps. I analysed the data, with help from Kelsey Byers and Richard Merrill. Anna Orteu did the RNA sequencing analysis. I carried out the functional analysis experiments with help from Ian Warren. Daiane and Stefan Schulz carried out chemical synthesis and analysis. Ana Pinharanda and John Davey carried out the assembly and annotation of the *H. cydno* genome. I wrote the manuscript and Richard Merrill, Owen McMillan, Kelsey Byers and Chris Jiggins contributed to the chapter concept. All authors gave feedback on the manuscript.

Chapter 5 is available as a pre-print:

Darragh, K., A. Orteu, K. J. R. P. Byers, D. Szczerbowski, I. A. Warren, P. Rastas, A. L. Pinharanda, J. Davey, S. F. Garza, D. A. Almeida, R. M. Merrill, W. O. McMillan, S. Schulz, and C. D. Jiggins. 2019. A novel terpene synthase produces an anti-aphrodisiac pheromone in the butterfly *Heliconius melpomene*. bioRxiv 779678.

Some of my work included in the appendix to this thesis is part of a collaboration. It is available as a pre-print:

Byers, K. J. R. P., K. Darragh, J. Musgrove, D. A. Almeida, S. F. Garza, I. A. Warren, P. Rastas, M. Kučka, Y. F. Chan, R. M. Merrill, S. Schulz, W. O. McMillan, and C. D. Jiggins. 2019. A major locus controls a biologically active pheromone component in *Heliconius melpomene*. bioRxiv 739037. doi: 10.1101/739037.

Contents

Chapter 1

Introduction

Olfaction is considered to be the most ancient and widespread sensory system (Ache and Young 2005; Wyatt 2014). Many groups, such as bacteria, plants and animals, use chemical cues to interact both with their environment and with other organisms (Wyatt 2014; Amo and Bonadonna 2018). In insects, chemical signalling is important for a range of behaviours such as aggregation (Wertheim et al. 2005), trail recruitment (Czaczkes et al. 2015), sex (Ali and Morgan 1990; Wyatt 2014), and interactions with plants (Schiestl 2010). These cues can be short-range, such as cuticular hydrocarbons, or long-range as in the case of volatile compounds released by female moths (Löfstedt 1993; Mas and Jallon 2005; Smadja and Butlin 2008; Wicker-Thomas 2011; Grillet et al. 2012).

Sex pheromones are chemical signals that mediate intraspecific communication important for mating (Wyatt 2003, 2014). The first sex pheromone to be chemically characterised was that of the moth *Bombyx mori* (Butenandt et al. 1959). Calling female moths release pheromones to attract mates (Itagaki and Conner 1988). These female pheromones allow males to find mates over long distances (up to a few miles), especially important in night-flying moths. Furthermore, they play an important role in mate recognition, and can lead to reproductive isolation and speciation if signals and preferences diverge between populations (Schneider 1992; Johansson and Jones 2007; Smadja and Butlin 2008; Saveer et al. 2014).

In many moth species, males also produce pheromones. These male pheromones act at a shorter range and are important in courtship and mating decisions (Phelan and Baker 1987; Schneider 1992; Lassance and Löfstedt 2009). Whilst male pheromones are not as well-studied as their female counterparts, a wide range of compound types, scent-disseminating male structures and behavioural roles have been described (Conner and

Iyengar 2016). They play a role in female acceptance, and can convey information about mate quality and species identity (Conner and Iyengar 2016).

In contrast to moths which mostly fly at night, most butterflies are diurnal. It is therefore thought that butterflies rely more heavily on vision to find mates (Vane-Wright and Boppré 1993). This agrees with the fact that long-distance female pheromones have not been described in diurnal butterflies. Similar to moths, however, male-emitted pheromones play a role in close-range courtship interactions (Schneider 1992; Vane-Wright and Boppré 1993). Within true butterflies, male pheromones have been described in Papilionidae, Pieridae, Lycaenidae and Nymphalidae (Meinwald et al. 1969; Pliske and Eisner 1969; Lundgren and Bergström 1975; Grula et al. 1980; Nishida et al. 1996; Ômura et al. 2001; Ômura and Honda 2005; Yildizhan et al. 2009; Wang et al. 2014; Ômura and Yotsuzuka 2015; Okumura et al. 2016; Pinzari et al. 2018).

Male pheromones are released from scent organs, which are derived from modified scales or hairs, found on the wings, abdominal organs, or appendages (Vane-Wright 1972). Different varieties, such as hair pencils and patches of specialised scales, can be found within the same family of butterflies (Hall and Harvey 2002). The chemical bouquets released by male scent organs often consist of multiple types of compounds. Macrolides, alkanes, alkenes, alcohols, ketones, aldehydes, and terpenoids can all be found within a single species (Schulz et al. 1993; Yildizhan et al. 2009). Some of these compounds, however, are not part of the pheromone but rather carrying out other functions such as matrix and carrier molecules, as well as cuticular components.

Information conveyed by sex pheromones in Lepidoptera

Male pheromones are important for courtship and stimulate female acceptance behaviours in some butterflies and moths (Brower and Jones 1965; Meinwald et al. 1969; Pliske and Eisner 1969; Grula et al. 1980; Phelan and Baker 1990; Nishida et al. 1996; Hillier and Vickers 2004). What is often unclear, however, is whether the compounds are merely acting as aphrodisiacs to increase mate acceptance, or if they convey information such as species identity or male quality.

Female-emitted moth pheromones play a role in species recognition and

reproductive isolation (Löfstedt et al. 1991; Smadja and Butlin 2008; Wicker-Thomas 2011). This could also be the case for male pheromones. Species which share a host plant with closely related species are more prone to mating mistakes, and also more likely to have male scent structures (Phelan and Baker 1987). Furthermore, patterns of character displacement in *Bicyclus* suggest that male sex pheromones contribute to reproductive isolation (Bacquet et al. 2015). If pheromones play a role in species recognition, this would predict stabilising selection towards a species stereotype (Löfstedt 1993; Johansson and Jones 2007). Species-specificity of male chemical profiles has been seen in both *Pyrgus* and *Bicyclus* (Hernández-Roldán et al. 2014; Bacquet et al. 2015). Finally, behavioural trials have also shown that male pheromones also contribute to reproductive isolation in both moths and butterflies (Grula et al. 1980; Lassance and Löfstedt 2009; Dopman et al. 2010; Hillier and Vickers 2011).

Not all male pheromones have species-specific compounds (Conner and Iyengar 2016). For instance, many male arctiid moths use the pheromones danaidal and hyroxydanaidal, derived from pyrrolizidine alkaloids (Schulz 2009). It is likely that instead of signalling species identity, these pheromones are related to mate quality (Conner and Iyengar 2016). This can be reflective of both direct benefits, such as resources, and indirect benefits, such as genetic quality, for the female (Andersson 1994). For pheromones to be an honest signal of quality, we expect them to be costly to produce, ensuring that only the best quality males can display the preferred pheromonal phenotype (Zahavi 1975, 1977), and indeed the cost of pheromone production has already been demonstrated in female moths (Harari et al. 2011).

Male pheromones can be condition-dependent, signalling mate quality. In the butterfly *Bicyclus anynana*, male sex pheromone production is affected by both inbreeding and male age (Nieberding et al. 2012; van Bergen et al. 2013). Females prefer to mate with mid-aged males rather than younger males, which correlates with changes in pheromone composition. Older males may be of higher quality as they have already proved their viability by surviving for longer (Brooks and Kemp 2001). Male tobacco moths, *Ephestia elutella*, that are large also have a higher mating success, sire larger offspring, and produce a higher amount of male pheromone (Phelan and Baker 1986). In these cases it seems that male pheromones are signalling indirect benefits to females

reflecting “good genes” which increase the survivorship or mating success of offspring (Andersson 1986, 1994).

Male sex pheromones can also signal the provision of a direct resource to the female. This is particularly common when the resource and pheromone are biochemically related. Compounds sequestered from host plants can act as sex pheromones, or pheromonal precursors, as well as protective compounds (Landolt and Phillips 1997). One of the best studied examples of this is the sequestration of pyrrolizidine alkaloids (PAs) by the arctiid moth *Uthetheisa ornatrix*. The male pheromone of this species, hydroxydanaidal, is derived from the sequestered PAs (Conner et al. 1981; Eisner and Meinwald 2003). PAs are transferred to the female during mating and chemically protect the eggs (Dussourd et al. 1988, 1991). The amount of male pheromone produced is an honest signal of the amount of stored alkaloid, thus signalling a direct benefit to the female (Eisner and Meinwald 1995; Iyengar et al. 2001).

Macroevolution of chemical profiles

Studying the broad-scale phylogenetic and geographic patterns of chemical profiles can also help us understand the role of chemical signalling. One hypothesis could be that chemical profiles would be species-specific with little variation if important for species recognition. On the other hand, we might predict high levels of variation in profiles, not necessarily following species boundaries, if profiles signal mate quality. However, species identity and mate quality are not mutually exclusive cues. Signals used in intraspecific mate assessment can diverge resulting in behavioural isolation (Ryan and Rand 1993; Johansson and Jones 2007; Smadja and Butlin 2008; Mendelson and Shaw 2012). Even if pheromones are important for behavioural isolation, they can exhibit variation both within and between populations of the same species, which could be important for mate quality assessment (Carde and Allison, 2016). We would expect to find at least certain species-specific characteristics to remain constant across the species range (Ferreira and Ferguson 2002; Benedict and Bowie 2009; McPeek et al. 2011; Weber et al. 2016). Studying patterns of chemical variation both inter- and intra-specifically can thus help to understand the role of pheromones in different systems.

Furthermore, we can attempt to understand not only the current role of chemical signalling, but also how profiles are evolving. Signal divergence can be driven by multiple evolutionary forces, including both neutral and adaptive forces, such as selection and drift (Leonhardt, Rasmussen, & Schmitt, 2013; Sun et al., 2013). One way to investigate this is through phylogenetic studies, whereby strong phylogenetic signal suggests that profiles are evolving under genetic drift, at the same rate as species divergence. However, this requires a large sample size across a well-described phylogeny. Another similar approach is to attempt to correlate genetic distance with phenotypic distance, again with a strong positive correlation being consistent with a role for genetic drift (Irwin, Thimgan, & Irwin, 2008). In this case, a lack of correlation suggests that selection is driving profile evolution (Campbell et al., 2010; Conrad, Paxton, Assum, & Ayasse, 2018; Hankison & Ptacek, 2008; Mullen, Vignieri, Gore, & Hoekstra, 2009). These studies can be carried out using different populations across the range of a species which are related to greater or lesser extents and exhibit varying phenotypes. Macroevolutionary studies prove an exciting and complementary approach to studies focusing on one species in detail (Weber et al. 2016).

Heliconius *butterflies as a study system*

Heliconius are a group of Neotropical butterflies that have been studied extensively to understand speciation and adaptation (Merrill et al. 2015). They first caught the attention of Henry Walter Bates due to their warning colour patterns which change with geographic location (Bates 1862). He saw this as an example of natural selection in action, in support of Darwin's theory of evolution by natural selection (Darwin 1863). Bates also described convergence in colour pattern between species in the same location, resulting in the formation of mimicry rings in the Amazon, and led him to formulate the theory of mimicry (Bates 1862; Jiggins 2017).

The warning coloration of *Heliconius* advertise their unpalatability to predators (Brower et al. 1963; Chai and Srygley 1990; Srygley and Chai 1990). Distastefulness is due to the presence of cyanogenic compounds in adult butterflies. *Heliconius* larvae can sequester cyanogenic compounds from the *Passiflora* on which they feed, and also synthesise compounds *de novo* (Nahrstedt and Davis 1985; Engler-Chaouat and Gilbert 2006; Castro et al. 2018). As adults, they are unique within butterflies as the only genus

that exhibit pollen feeding (Gilbert 1972). They process pollen on their proboscis to extract amino acids which facilitates a long reproductive lifespan (Dunlap-Pianka et al. 1977). It has been proposed that reduced dependence on larval resources for reproduction, as eggs are made from adult resources, allows larval resources to be directed towards investment in defensive compounds and rapid development (Cardoso and Gilbert 2013). Both larval and adult diet therefore play an important role in the ecology of *Heliconius*.

Over the past 150 years, the wing patterns of *Heliconius* butterflies have been the focus of studies on mimicry and ecology as well as more recent work on genetics and development (Merrill et al. 2015; Jiggins 2017). Due to the day-flying nature of butterflies, warning colours play a key role in mate-finding and recognition. Male *Heliconius* exhibit a preference for females of their own colour pattern (Jiggins et al. 2001; Kronforst et al. 2006; Melo et al. 2009; Muñoz et al. 2010; Merrill et al. 2011a,b, 2014; Sánchez et al. 2015). In many cases, closely related species pairs differ in their warning colour pattern, suggesting that mimicry shifts could lead to reproductive isolation and speciation (Jiggins et al. 2001; Jiggins 2008).

Colour pattern and male mate preference divergence cannot explain all cases of reproductive isolation in *Heliconius*. Some closely related species are also co-mimics (Brower 1996; Giraldo et al. 2008; Mérot et al. 2013). These species exist in sympatry and exhibit strong assortative mating, despite sharing a colour pattern, implicating other cues in mate choice (Giraldo et al. 2008). Even in cases where species do differ in colour pattern, isolation due to colour pattern preference alone is not as strong as assortative mating, suggesting further barriers are required (Jiggins 2008). These earlier studies suggest a combination of chemical and visual cues are important in *Heliconius*. The importance of multiple cues for mate choice is now widely appreciated in the literature (Candolin 2003; Papke et al. 2007; Girard et al. 2015; Wegehaupt and Wagner 2017).

Female mate choice in Heliconius

Heliconius males exhibit strong visual preferences, and male choice acts early in courtship and so can be an important isolating barrier between species (Merrill et al.

2019). Although females have traditionally been thought of as the 'choosy' sex due to their higher investment in offspring (Trivers 1972), in *Heliconius* we expect male choice as investment by males is also high. Males donate a nutrient-rich spermatophore during mating which is used in both female somatic tissue and in eggs (Boggs and Gilbert 1979; Boggs 1981, 1990). Male choosiness is common in Lepidoptera where male investment is high (Bonduriansky 2001). Furthermore, in some species, males mate with females as they emerge from the pupa, when females are unable to exert choice (Edwards 1881; Gilbert 1991; Deinert et al. 1994; Thurman et al. 2018). At least in *H. erato*, however, it seems that adult mating may be more common in the wild than previously thought (Thurman et al. 2018). When butterflies mate as adults, forced copulation is rare in Lepidoptera (Wiklund and Forsberg 1986; Forsberg and Wiklund 1989).

In contrast there has been much less investigation of female mate choice in *Heliconius*. Despite male investment, the operational sex ratio (OSR) suggests that females should also be choosy (Emlen and Oring 1977; Clutton-Brock and Vincent 1991; Clutton-Brock and Parker 1992). The OSR can be affected, for example, if one sex spends more time rearing offspring, or one sex only mates once. The sex that is less 'available' overall is in high demand and so will be more choosy. Re-mating is rare in female *Heliconius*, perhaps because pollen-feeding reduces the nutritional benefit of multiple matings (Walters et al. 2012; Merrill et al. 2015). The mating decision for females is, therefore, proportionally more important than for males. We also need to look beyond the dichotomy of male vs female choice as it is likely that both males and females can choose whether or not to mate, in other words, there is mutual mate choice (Cunningham and Birkhead 1998; Bergstrom and Real 2000).

There is considerable indirect evidence for female choice in *Heliconius*. Males may court other species or races, but fail to achieve mating, suggesting a role for female rejection (Jiggins et al. 2004; Mérot et al. 2015). More recently, attempts have been made to directly measure female preference. Females of *H. numata* exhibit a disassortative mating preference, contributing to the maintenance of an intraspecific wing pattern polymorphism (Chouteau et al. 2017). Female mate choice also contributes to reproductive isolation between species of *Heliconius* (Southcott and Kronforst 2018).

Furthermore, *Heliconius* female rejection behaviours have been documented, including raising their abdomen and flattening their wings (Mallet 1986; Klein and Araújo 2010).

Chemical signalling in Heliconius

It has long been appreciated that odours, as well as visual cues, are used in mating (Crane 1955), but most work in *Heliconius* has emphasised visual cues in both male and female choice (Jiggins et al. 2001, 2004; Chouteau et al. 2017; Southcott and Kronforst 2018). Indeed, *Heliconius* have increased visual complexity due to the duplication of a UV opsin (Briscoe et al. 2010). Nonetheless, genome sequencing revealed a wealth of chemosensory genes (Heliconius Genome Consortium 2012) that are likely to be involved in various processes, including host plant finding and mating (Briscoe et al. 2013; van Schooten et al. 2016). For example, chemical cues are used by males of pupal mating species to locate mates. Caterpillars cause leaf tissue damage which increases the release of some compounds, which are then used by male *H. charithonia* to find mates (Estrada and Gilbert 2010). They can then also determine the sex of the pupae using chemical cues (Estrada et al. 2010). Males also seem to use chemical cues when courting adult butterflies. *Heliconius erato* males can distinguish between conspecific and co-mimic wings only when wings have not been washed in hexane (Estrada and Jiggins 2008).

Early studies of *Heliconius* often note the strong smell emitted from the genital region by both males and females when captured (Müller 1912a; Poulton 1925). These scents are described as a repulsive odour, related to protection from predators (Eltringham 1925). Due to the proposed anti-predation function, they are described as being part of the mimicry phenotype (Collenette 1929). Aposematic warning signals may therefore be multimodal in nature, with chemical compounds aiding predator recognition (Rothschild 1961; Rojas et al. 2018). There is, however, no direct evidence for a role in predator deterrence.

Whilst it is unclear if genital compounds act as anti-predation cues, we now know that they are certainly important for chemical mate guarding. Anti-aphrodisiac compounds are transferred from males to females during mating to prevent re-mating by making the female unattractive to males, and were first described in Lepidoptera in the

1970s in *H. erato* (Gilbert 1976; Malouines 2017). The compound acting as an anti-aphrodisiac in *H. melpomene* has since been identified as (*E*)-β-ocimene (Schulz et al. 2008). As well as potentially leading to sexual conflict over reduced re-mating, anti-aphrodisiacs could act as honest signals of receptivity, reducing harassment while females are unreceptive to further matings (Estrada et al. 2011). Compounds found in male genitalia, which potentially act as anti-aphrodisiacs, are diverse, with both type and number of compounds varying between species (Estrada et al. 2011). The role of these compounds in female choice remains unclear.

As well as the repulsive compounds found in the genital region, Müller described 'scent scales' in male *Heliconius*, thought to be aphrodisiac in nature (Müller 1912b). 'Scent scales' have a brush-like structure to aid the emission of sex pheromones (Müller 1912b; Eltringham 1925; Barth and Barth 1952; Emsley 1963). The region containing these scales is known as the androconia and is found in the silvery overlapping region of the wing (Emsley 1963). During courtship, males expose the androconia, whilst rapidly moving both their fore- and hindwings, presumably to expose the female to the compounds which are released (Crane 1955). Exposure of the androconial region occurs in all successful matings of *H. erato* (Klein and Araújo 2010).

Despite early recognition of the importance of chemical signalling in *Heliconius*, we know very little about male pheromones in this group. Recent efforts to characterise the chemistry of the androconial region have found evidence for species differences in composition (Mérot et al. 2015; Mann et al. 2017) . Furthermore, *H. melpomene* females prefer to mate with *H. melpomene* males which have been "perfumed" with a pheromone extract from a conspecific male when compared to a *H. timareta*-perfumed male (Mérot *et al.,* 2015). This "perfuming" experiment included both genital and androconial compounds and so it is unclear which had the effect on female choice. Whilst the role of chemical signaling and female choice in *Heliconius* are both relatively unexplored areas, it seems likely that female choice based on chemical cues is important for mate choice in the genus.

Where do pheromones come from?

In order to understand the information that is conveyed by pheromones we need to know their source. Many male pheromones are similar in structure to compounds found in plants (Baker 1989; McNeil and Delisle 1989; Birch et al. 1990). This suggests that males do not synthesise these compounds *de novo* and instead ingest precursors and alter them to produce pheromones, as in *Uthetheisa ornatrix* (Conner et al. 1981; Eisner and Meinwald 2003). If females already use volatile cues to find host plants, the evolution of similar male pheromones could take advantage of existing sensory bias (Baker 1989; Conner and Iyengar 2016).

Some Lepidoptera synthesise pheromones *de novo* (Clearwater 1975; McNeil and Delisle 1989; Conner and Iyengar 2016). These pheromones are often long-chain hydrocarbons, similar to those produced by female moths (Hillier and Vickers 2004; Nieberding et al. 2008; Lassance and Löfstedt 2009). It is thought that pre-existing biosynthetic pathways can therefore be co-opted by males to produce new pheromones (Conner and Iyengar 2016). Two key pheromone components in *Bicyclus anynana* are produced by conserved genes via a female moth-like biosynthetic route (Liénard et al. 2014). Furthermore, these pheromones are formed by modification of fatty acids, which are productions of primary metabolism (Groot et al. 2016), and thus may be easier to evolve than structurally novel pheromones.

The origin of chemical compounds found in the androconial and genital regions of *Heliconius* is generally unclear. Of the number of androconial compounds described, around 70% originate from fatty acid biosynthesis, similar to *Bicyclus*, and are thought to be biosynthesised by the butterflies (Bacquet et al. 2015; Mann et al. 2017). The anti-aphrodisiac compound of *H. melpomene*, (*E*)-β-ocimene, can be synthesised by males (Schulz et al. 2008). (*E*)-β-Ocimene is also a common floral scent (Farré-Armengol et al. 2017), and so even compounds which look plant-like in structure can potentially be synthesised by Lepidoptera, preventing us from inferring compound origin (sequestration or biosynthesis) from chemical structure.

Genetics and evolution of pheromone production

Ultimately, to understand how pheromones evolve we need to know their genetic basis. Whilst the pheromone composition of many species has been described, the genetic changes underlying chemical diversity is often not known (El-Sayed 2007; Symonds and Elgar 2008; Groot et al. 2016). To date, the genetics of butterfly pheromone production has only been studied in *Bicyclus anynana* (Liénard et al. 2014). Classic debates in trait evolution include whether gradual or saltational changes (small or large effect mutations) are more important (Baker 2002; Symonds and Elgar 2008), and the importance of coding vs. regulatory changes (Roelofs et al. 2002; Lassance et al. 2013).

Determining the genetic basis of pheromone production can also help us to understand convergence, the evolution of the same phenotype between different species. Convergence has been studied in a diverse range of systems, and can be due to differing molecular mechanisms (Stern 2013). It can be the result of independent genetic mutations in the different species, shared ancestry or hybridisation, or even horizontal gene transfer (Stern 2013). To know which of these scenarios has occurred we need to know the genetic basis of the trait of interest in both lineages.

Plants and insects often use the same or very similar compounds for attraction and defence (Schiestl 2010). In most cases where both plants and insects are biosynthesising the same compound the genetic basis of convergence is unclear (Beran et al. 2019). Some of these biosynthetic pathways, such terpene production, are better described in plants and we are only beginning to investigate whether the same pathways are used by insects. One possibility is that the same biosynthetic machinery is used due to horizontal transfer of genes between plants and insects. This has not been recorded for pheromone biosynthesis but could be possible as aphid carotenoid biosynthesis evolved from genes transferred from fungi (Moran and Jarvik 2010). Nonetheless, it seems most likely that biosynthetic pathways evolved independently in plants and insects (Beran et al. 2019).

An interesting example where we can investigate convergent evolution is the production of (*E*)-β-ocimene by *H. melpomene* males. (*E*)-β-Ocimene belongs to the

largest and most structurally diverse group of natural product compounds, the terpenes (Gershenzon and Dudareva 2007). Until recently, terpene synthase (TPS) genes were only known from plants and fungi among the eukaryotes, with the suggestion that insects could not synthesise terpenes *de novo*, instead deriving them from plants (Chen et al. 2016). Recently, insect TPSs, which are unrelated to plant TPSs, have been described in Hemiptera and Coleoptera, suggesting an independent evolutionary origin (Gilg et al. 2005, 2009; Beran et al. 2016, 2019; Lancaster et al. 2018, 2019). β-Ocimene synthases have been described in plants but not animals, allowing us to investigate convergence between plants and insects (Farré-Armengol et al. 2017).

It is still unclear if the evolution of TPS activity has occurred only once in insects or multiple times (Beran et al. 2019; Lancaster et al. 2019). Interestingly, ocimene is used as a recruitment pheromone by *Bombus terrestris* and a larval pheromone by *Apis mellifera* (Granero et al. 2005; Maisonnasse et al. 2010). It is not known if biosynthesis in Hymenoptera has an independent evolutionary origin to that in Lepidoptera. Determining the genetic basis of (*E*)-β-ocimene production in *Heliconius* will shed light not only on the evolutionary origin of TPSs activity in insects, but also the convergent use of the same compounds for different function between insect lineages.

Many of the enzymes involved in pheromone production are members of large protein families (Groot et al. 2016). These families are often characterised by gene duplication events, as well as the presence of pseudogenes (Nei et al. 1997; Eirín-López et al. 2012). Moth desaturases seem to follow a birth-and-death model of multi-gene family evolution (Roelofs and Rooney 2003). Gene families expand and contract through gene duplication and gene deletion (Nei and Rooney 2005). Following duplication, genes can be lost by deterioration to pseudogenes, gain new functions through neofunctionalization, or retain ancestral functions through subfunctionalisation (Wagner 1998; Force et al. 1999; Rastogi and Liberles 2005; Assis and Bachtrog 2013).

Fatty acyl reductases and desaturases involved in moth pheromone production belong to large protein families. Subfunctionalisation following gene duplication could explain pheromone differences due to desaturases between *Ostrina* moth species (Roelofs et al. 2002). Coding differences between fatty acyl reductases can alter substrate

preference between the enzymes of different species, and so enzymes can evolve new functional properties as well as partitioning ancestral ones (Lassance et al. 2013). Plant TPSs can be categorised into seven subfamilies, highlighting the diversity of these enzymes (Chen et al. 2011). Gene duplication and neofunctionalization play an important role in their evolution, with closely-related enzymes able to produce different compounds (Nagegowda et al. 2008; Chen et al. 2011). To date, lineage-specific duplications have only been described in one family of TPS genes in insects (Beran et al. 2016), likely due to the limited number of insect TPS genes described. Investigating the genetic basis of (*E*)-β-ocimene in *Heliconius* butterflies will allow us to ask how TPS genes have evolved in this group.

Challenges in chemical ecology

The first step in understanding the chemical ecology of a species is usually the observation of a behavior which suggests chemical signaling is important. Once this is established, the potential pheromone components need to be identified. Chemical extracts are analysed by GC-MS and the resulting spectra compared to known compound libraries and chemical standards. Synthesis may be required to identify more unusual components of the blend. The pheromone blend of hundreds of species of Lepidoptera can be found on Pherobase, a database for pheromone components (El-Sayed 2007). In many cases, however, it is not clear which components of the blend act as the pheromone.

To identify the active components of the blend we need a bioassay. Behavioural assays of female pheromones are relatively straight forward as male attraction to compounds can be measured in wind tunnels. The response of females to male pheromones is generally harder to measure, especially without a detailed understanding of courtship in the species of interest. This is certainly true in *Heliconius*, where female behaviour has rarely been studied (but see: Klein and Araújo 2010; Chouteau et al. 2017; Southcott and Kronforst 2018).

In these cases where behavioural trials are difficult or time-consuming, electroantennograms can be used as a complementary approach. This technique

measures the electrophysiological response of antennae to a compound or mix of compounds. Whilst this does not determine whether a response is positive or negative, it at least can show which components of a blend may be biologically active. In *Heliconius*, this has been used to determine an active component of the androconial bouquet of *H. melpomene*: octadecanal (Byers et al. 2019). A further technique, gas-chromatography coupled with electroantennographic detection (GC-EAD), involves separation of a blend whilst measuring antennal response, thus allowing identification of all active components in the blend (Arn et al. 1975). GC-EAD has been used to investigate responses of *H. melpomene* to floral scents, where a strong antennal response to the plant volatile (*E*)-β-ocimene was detected (Andersson and Dobson 2003). (*E*)-β-Ocimene is also the anti-aphrodisiac of *H. melpomene*, which has been confirmed behaviourally (Schulz et al. 2008).

In this thesis I propose a complementary statistical approach to bioassays. We expect pheromones to experience stabilizing selection towards a species stereotype, especially if they play a role in species recognition (Löfstedt 1993; Johansson and Jones 2007). These species-specific characteristics are therefore likely to remain constant across the species range (Ferreira and Ferguson 2002; Benedict and Bowie 2009; McPeek et al. 2011; Weber et al. 2016). We can therefore look at patterns of intraspecific variation across a geographic range to identify invariant components of the blend, found in all individuals. Furthermore, we can try to identify compounds or combinations of compounds which uniquely identify a species, in other words, they are found in all individuals of a certain species but not in individuals from other species in the study. This "indicator" approach seems a promising first step in pheromone identification. Compounds identified by statistical methods or electrophysiology must then be tested behaviourally, but at least these initial steps narrow the list of candidates significantly.

Outline of thesis

In this thesis I investigate the role of chemical signalling in mate choice of *Heliconius* butterflies. I investigate the effect of diet on pheromone production, offering insight into the information conveyed by these signals. I use a broad geographic study to look for patterns of evolution both within and between species, as well as identifying

species-specific compounds, which are candidates for species recognition components. I also study the genetic basis for the biosynthesis of the terpene (*E*)-β-ocimene, which is used by *H. melpomene* as an anti-aphrodisiac, providing insights into convergent evolution both between plants and insects, and also between different insect lineages. Together, these results provide insights into the chemistry, ecology, evolution, and genetics of an understudied signal in a butterfly with a long history of visual studies.

Chapter 2

Male sex pheromone components in *Heliconius* butterflies released by the androconia affect female choice

Abstract

Sex specific pheromones are known to play an important role in butterfly courtship, and may influence both individual reproductive success and reproductive isolation between species. Extensive ecological, behavioural and genetic studies of *Heliconius* butterflies have made a substantial contribution to our understanding of speciation. Male pheromones, although long suspected to play an important role, have received relatively little attention in this genus. Here, we combine morphological, chemical and behavioural analyses of male pheromones in the Neotropical butterfly *Heliconius melpomene*. First, we identify putative androconia that are specialized brush-like scales that lie within the shiny grey region of the male hindwing. We then describe putative male sex pheromone compounds, which are largely confined to the androconial region of the hindwing of mature males, but are absent in immature males and females. Finally, behavioural choice experiments reveal that females of *H. melpomene*, *H. erato* and *H. timareta* strongly discriminate against conspecific males which have their androconial region experimentally blocked. As well as demonstrating the importance of chemical signalling for female mate choice in *Heliconius* butterflies, the results describe structures involved in release of the pheromone and a list of potential male sex pheromone compounds.

Introduction

Sex pheromones are species-specific blends of chemical compounds that mediate intraspecific communication between males and females (Wyatt 2003, 2014). Among insects, pheromone communication can involve a single chemical, but often relies on a complex combination of multiple chemical components (Grillet et al. 2006; Nieberding et al. 2008; Symonds et al. 2012). This chemical complexity provides the potential to convey sophisticated information, such as the inbreeding status of the emitter (Ando et al. 2004; van Bergen et al. 2013; Menzel et al. 2016), mate quality (Dussourd et al. 1991; Ruther et al. 2009), and species identity (Danci et al. 2006; Saveer et al. 2014). Perhaps the best studied insect sex pheromones are those produced by female moths to attract mating partners, often over long distances (Löfstedt 1993; Smadja and Butlin 2008). However, male insects also produce sex pheromones (Eggert and Müller 1997; Kock et al. 2007; Ruther et al. 2009; Meinwald et al. 1969), and chemical signalling can occur over short distances (Nishida et al. 1996; Mas and Jallon 2005; Smadja and Butlin 2008; Wicker-Thomas 2011; Grillet et al. 2012).

Sex pheromones can play a key role in determining the reproductive success of individuals within a species, and may also result in reproductive isolation between species if signals diverge (Johansson and Jones 2007; Smadja and Butlin 2008; Wyatt 2014). Within Lepidoptera, the importance of chemical signalling in mate choice and speciation is well established among moth species (Phelan and Baker 1987; Löfstedt 1993; Bethenod et al. 2004; Dopman et al. 2010; Saveer et al. 2014). Most moths fly at night, when visual signalling is unlikely to be as effective in attracting mates. In contrast, butterflies are mostly diurnal and visual signals are usually important for initial mate attraction (Vane-Wright and Boppré 1993). However, chemical signals can play other roles in butterfly mate choice, with evidence that close-range courtship interactions often involve pheromones emitted by males, in contrast to the long-distance signalling with female-emitted pheromones more commonly observed in moths (Vane-Wright and Boppré 1993). Acceptance behaviour in the queen butterfly *Danaus berenice*, for example, is regulated by a dihydropyrrolizine alkaloid released by the male during courtship (Brower and Jones 1965; Meinwald et al. 1969; Pliske and Eisner 1969). Another danaine butterfly,

Idea leuconoe, displays brush-like structures, called 'hair-pencils', emitting a mixture of volatiles during courtship, which when applied to dummy males elicits an acceptance posture in females (Nishida et al. 1996). *Pieris rapae* and *P. brassicae* both use macrocyclic lactones as a pheromone to induce acceptance in females (Yildizhan et al. 2009). Finally, in *Bicyclus anynana* males with reduced amounts of male sex pheromone have decreased mating success, implying a direct involvement in reproductive fitness (Nieberding et al. 2008, 2012).

Here we focus on the potential role of male pheromones in *Heliconius* butterflies. *Heliconius* is a diverse Neotropical genus, which has been extensively studied in the context of adaptation and speciation (Jiggins 2008; Supple et al. 2014; Merrill et al. 2015). These butterflies are well known for Müllerian mimicry, in which unrelated species converge on the same warning signal to more efficiently advertise their unpalatability to predators. Closely related *Heliconius* taxa, however, often differ in colour pattern and divergent selection acting on warning patterns is believed to play an important role in speciation (Bates 1862; Jiggins et al. 2001; Merrill et al. 2011b).

Male *Heliconius* display conspicuous courtship behaviours likely because the availability of receptive females in nature is limited. Female re-mating is a rare event in *Heliconius* (Walters et al. 2012) and males must compete to find virgin females within a visually complex environment (Merrill et al. 2015). In addition, males donate a nutrient-rich spermatophore during mating (Boggs and Gilbert 1979; Boggs 1981) which, together with costs associated with extended copulation, will select for discrimination against less suitable mates in both sexes. A combination of colour (hue) and movement stimulates courtship by *Heliconius* males (Crane 1955). More recently, it has repeatedly been shown across multiple *Heliconius* species that males are more attracted to their own warning pattern than that of closely related taxa (Jiggins et al. 2001, 2004; Kronforst et al. 2006; Melo et al. 2009; Muñoz et al. 2010; Merrill et al. 2011a,b, 2014; Finkbeiner et al. 2014; Sánchez et al. 2015). In addition to colour pattern, male *Heliconius* also use chemical signals to locate and determine the suitability of potential mates. This includes the use of green leaf volatiles during mate searching. Six-carbon alcohols and acetates are released by host plants in larger amounts after leaf tissue damage caused by caterpillars, which adult males of the pupal mating species *H. charithonia* then use to find potential mates

(Estrada and Gilbert 2010). Once males find pupae they also use chemical cues to determine sex (Estrada et al. 2010). Supporting a further role for chemical signals, *Heliconius erato* males distinguish between wings dissected from conspecific and local *H. melpomene* females that are virtually identical in wing pattern, but this effect disappears after wings have been washed in hexane (Estrada and Jiggins 2008).

As well as attraction, chemicals can also be involved in repulsion. Males are repelled by a strong odour released by previously mated females (Gilbert 1976). This 'anti-aphrodisiac' is produced by males soon after eclosion and is then transferred during copulation (Schulz et al. 2008). The abdominal glands of male *H. melpomene*, for example, contain a complex chemical bouquet consisting of the volatile compound (*E*)-β-ocimene together with some trace components and esters of common C_{16} – and C_{18} – fatty acids with alcohols, where β-ocimene acts as the main anti-aphrodisiac component (Schulz et al. 2008). This anti-aphrodisiac effect occurs in several *Heliconius* species, which show species-specific patterns of scent gland constituents (Gilbert 1976; Estrada et al. 2011).

Despite the focus on male mate choice, analysis of courtship in *Heliconius* has shown that females can exhibit rejection behaviours, such as raising their abdomen and flattening their wings (Mallet 1986; Klein and Araújo 2010). There are also a number of observations that indicate a role for chemical recognition in female mate choice. *Heliconius erato* males separate their wings during courtship to reveal the silvery overlap region, suggested to be involved in the distribution of pheromones. This behaviour, described as androconial exposition, occurs in every courtship that results in mating, suggesting that pheromones influence the female response (Klein and Araújo 2010). Additionally, direct evidence that *Heliconius* females use chemical signals to distinguish conspecific males comes from studies of the closely related species *H. timareta* and *H. melpomene*, which share the same warning patterns in Peru (Mérot et al. 2015). Males experimentally treated with abdominal scent glands and wing extracts of heterospecifics show a reduced probability of mating. Chemical analysis of both abdominal gland and whole wings provides evidence for qualitative and quantitative differences in the chemical signatures between these closely related species (Mérot et al. 2015).

Here, we investigate the role of chemical signalling in female mate choice in *Heliconius* at three levels. First, we investigate morphological structures potentially associated with pheromone production. In butterflies, a variety of species-specific structures including brushes, fans, and differentiated scales on wings, legs or abdomen are used to expose pheromones produced in associated glands (Wyatt 2003; Nieberding et al. 2008). In particular, male-specific scent glands, termed androconia, are common across the Lepidoptera. In male *Heliconius*, a patch of shiny grey scales is present on the overlapping region of the hind and forewing (Fig. 1). The observed sexual dimorphism in this trait suggests that these are androconia, and may be associated with a male sex pheromone (Emsley 1963). Furthermore, earlier authors have identified brush-like scales in the hindwing androconial region that are the putative site for pheromone production and emission (Müller 1912b; Barth 1952). Here we investigate the structure of these scales using scanning electron microscopy. Second, we complement recently published chemical analysis of whole *H. melpomene* wings (Mérot et al. 2015) by dissecting wing regions to identify those associated with the production of compounds and identify the potential male sex pheromone compounds isolated from this region. As *H. melpomene rosina* was available for more extensive behavioural experiments at the insectaries in Panama, we focused further work on this population, including repeating chemical analyses to ensure that the sexual dimorphism was also present. Finally, we carry out mate choice experiments in *H. melpomene rosina*, *H. melpomene malleti, H. timareta florencia* and *H. erato demophoon* to test the importance of pheromones for female choice in *Heliconius*.

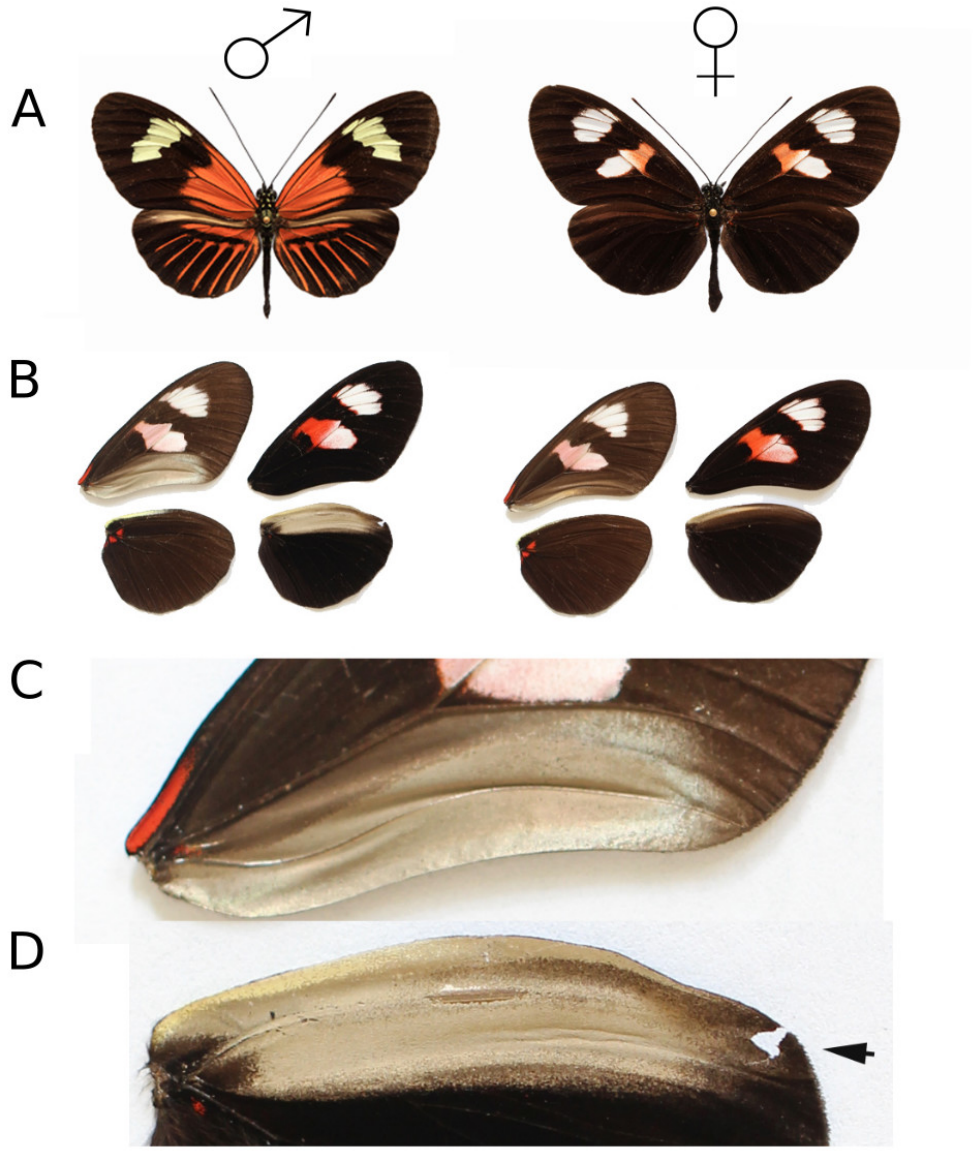


Figure 1: Heliconius melpomene *wings showing androconial dimorphism. (A)* H. melpomene malleti *(Ecuador sample, left) and* H. melpomene plesseni (*Ecuador sample, right). (B) Dissected wings from specimens of* H. melpomene plesseni *showing sexual dimorphism in the androconial region, with male (left) and female (right). For each sex, the left set of wings shows the ventral surface and the right set the dorsal surface. (C) Expanded view of the male forewing overlapping region. The pale grey-brown region was dissected for chemical analysis. (D) Expanded view of the male hindwing androconial region, with arrow highlight the vein Sc + R1. The pale grey-brown region was dissected for chemical analysis. The ventral side of the forewing is on the top and the dorsal side of the hindwing is on the bottom. The pale grey-brown region in the male wing was dissected for chemical analysis.*

Methods

Individuals used for morphological and chemical analyses were from an outbred stock of *Heliconius melpomene plesseni* and *Heliconius melpomene malleti* (sold as *H. m. aglaope*), maintained at the University of Cambridge insectaries (Fig. 1A). These two races are from the region of a hybrid zone in the eastern Andes of Ecuador, and showed considerable inter-racial hybridization in the stocks, so are treated here as a single population and referred to as the Ecuador samples. These stocks were established from individuals obtained from a commercial breeder (Stratford-Upon-Avon Butterfly Farm, Swans Nest, Stratford-Upon-Avon, CV37 7LS, UK: www.butterflyfarm.co.uk). Laboratory stocks were maintained on the larval food plants, *Passiflora menispermifolia* and *P. biflora*. Adult butterflies were fed on ~10% sucrose solution mixed with an amino acid supplement (Critical Care Formula®, Vetark Professional, Winchester, UK). Further chemical and behavioural analysis were carried out on the mimetic but distantly related *H. melpomene rosina* and *H. erato demophoon* reared at the Smithsonian Tropical Research Institute (STRI) facilities in Gamboa, Panama, and are referred to as the Panama samples. Both males and females of the Panama samples were from outbred stocks established from wild individuals collected in Gamboa (9˚7.4’ N, 79˚42.2’ W, elevation 60 m) within the nearby Soberania National Park, and San Lorenzo National Park (9°17’N, 79°58’W; elevation 130 m). Larvae were reared on *Passiflora williamsi* and *P. biflora*. Adult butterflies were provided with ~20% sugar solution with *Psychotria sp.*, *Gurania sp.*, and *Psiguiria sp.* as pollen sources. Finally, behavioural experiments were carried out in the mimetic and closely related species *Heliconius melpomene malleti* and *H. timareta florencia* reared at the insectaries of Universidad del Rosario (UR) in La Vega, Colombia. These stocks derived from wild caught individuals from Sucre, Caqueta (01°48'12" N, 75°39'19"W, elevation 1200 m). Larvae were reared on *Passiflora oerstedii* and adults were provided with *Psiguiria sp.* as pollen source and ~20% sugar solution.

Morphological analysis

The detailed morphology of androconial scales was determined using a Field Emission Scanning Electron Microscope. Three males and two females of *H. melpomene* from Ecuador were used for this analysis. The overlap grey scale region was dissected out

from both hind and forewings and attached to aluminium stubs with carbon tabs and subsequently coated with 20nm of gold using a Quorum/Emitech sputter coater. The gold-coated regions were then viewed in an FEI XL30 FEGSEM operated at 5kV. Images were recorded digitally using XL30 software at 500x magnification.

Characterization of potential male sex pheromone

Wing tissue from ten males (five newly emerged and five 10-day old) and five females (10-day old) from the Ecuador stock was collected between November 2011 and March 2012 for chemical analysis. Wings were dissected into four parts: forewing overlap, hindwing androconia, forewing rest and hindwing rest. The 'androconia' and 'overlap' regions corresponded to the grey-brown region (Fig. 1B, 1C and 1D), with rest corresponding to the remaining portion of the wing which is not overlapping. In females, a region corresponding in size and extent to the grey-brown region seen in males was dissected. The dissected sections were then immediately placed in 200µl hexane or dichloromethane in 2mL glass vials and allowed to soak for three hours. Initial analysis showed no major differences in extracted chemicals between hexane and dichloromethane extracts (data not shown). Therefore, the more polar dichloromethane was used in later analyses. Due to a larger available stock of *H. melpomene rosina* for behavioural experiments, hindwing androconial tissue was also then collected from 20 males and 11 females (both 10-12 days old) in Panama between February and July 2016. The tissue was soaked in 200µl dichloromethane in 2ml glass vials, with PTFE-coated caps, for one hour. The extraction time was shortened as this had no influence on the results (Fig S1). The solvent was then transferred to new vials and stored at -20°C. Samples were evaporated under ambient conditions at room temperature prior to analysis.

Mature male androconial extracts from the Ecuador stock were analysed by gas chromatography/mass spectrometry (GC/MS) using a Hewlett-Packard model 5975 mass-selective detector connected to a Hewlett-Packard GC model 7890A, and equipped with a Hewlett-Packard ALS 7683B autosampler. All other Ecuador extracts were analysed by comparison to the male androconial results. Extracts from the Panama stock were analysed by GC/MS using a Hewlett-Packard model 5977 mass-selective detector connected to a Hewlett-Packard GC model 7890B, and equipped with a Hewlett-Packard

ALS 7693 autosampler. HP-5MS fused silica capillary columns (Agilent, 30 m × 0.25 mm, 0.25 µm) were used in both GCs. In both cases, injection was performed in splitless mode (250°C injector temperature) with helium as the carrier gas (constant flow of 1.2 ml/min). The temperature programme started at 50°C, was held for 5 min, and then rose at a rate of 5°C/min to 320°C, before being held at 320°C for 5 minutes. Components were identified by comparison of mass spectra and gas chromatographic Kovats retention index with those of authentic reference samples and also by analysis of mass spectra. The double bond positions of unsaturated compounds were determined by derivatisation with dimethyl disulfide (Buser et al. 1983). To confirm chemical structures, alcohols were synthesised from the corresponding methyl esters by reduction according to established procedures (Becker and Beckert 1993 p. 570). The aldehydes were synthesised by oxidation of the respective alcohols (More and Finney 2002).

Compounds found in the extracts were quantified using gas chromatography with flame ionisation detection with a Hewlett-Packard GC model 7890A or 7890B equipped with a Hewlett-Packard ALS 7683B (Ecuador) or 7693 (Panama) autosampler. A BPX-5 fused silica capillary column (SGE, 25 m × 0.22 mm, 0.25 µm) was used in both cases. Injection was performed in splitless mode (250°C injector temperature) with hydrogen as the carrier gas (constant flow of 1.65 ml/min). The temperature programme started at 50°C, held for 5 min, and then rose to 320°C with a heating rate of 5°C/min. Pentadecyl acetate (10.1 ng) or (*Z*)-4-tridecenyl acetate (1 ng) were used as internal standard for Ecuador samples, and 2-tetradecylacetate (200ng) for Panama samples. Only compounds eluting earlier than hexacosane were considered for analysis. Later compounds were identified as cuticular hydrocarbons, 2,5-dialkyltetrahydrofurans, cholesterol and artefacts (*e.g.* phthalates or adipates). The variability in the late eluting cuticular hydrocarbons was low and did not show characteristic differences between samples.

For the Ecuador samples, groups were visualised as boxplots, due to the high frequency of absent compounds in the samples. We then used non-parametric Kruskal-Wallis to test for differences between the amounts of compounds present in different wing regions of mature males, and also between age and sex categories. This was followed up by Dunn post-hoc testing (Dinno 2017; Ogle 2017), with Bonferroni correction.

For the Panama samples, due to the higher sample size, and larger number of compounds identified we visualised the males and females as two groups using a non-metric multidimensional scaling (NMDS ordination, based on a Bray-curtis similarity matrix. We used the metaMDS function in the package vegan (Oksanen et al. 2017), with visualisation using the package ade4 (Dray and Dufour 2007). This was followed up with ANOSIM to compare differences between groups, and non-parametric Kruskal-Wallis tests to determine which compounds differed between sexes. All statistical analyses were performed with R version 3.3.1 (R Core Team 2018).

Behavioural experiments

To test female acceptance of male pheromones, behavioural experiments were conducted in insectaries at STRI, Gamboa, Panama between February and July 2016, and also in insectaries at UR in La Vega, Colombia between November 2015 and June 2016. One day old virgin females were presented with a control male and a 'pheromone blocked' male, both of which were at least ten days old. Males from Panama were treated with transparent nail varnish (Revlon Liquid Quick Dry containing cyclomethicone, isopropyl alcohol, ethylhexyl palmitate, mineral oil and fragrance) applied to wings, following Constanzo and Monteiro (Costanzo and Monteiro 2007). Males from Colombia were treated with transparent nail varnish (Vogue Fantastic containing butyl acetate, ethyl acetate, nitrocellulose, adipic acid, neopentyl glycol, trimellitic anhydride copolymer, isopropyl alcohol, acetyl tributyl citrate, stearalkonium bentonite, styrene, acrylates copolymer, silica benzophenone-1, calcium sodium borosilicate, synthetic fluorphlogopite, polyethylene terephthalate and polyurethane-11). Pheromone blocked males had the dorsal side of their hindwing androconia blocked, whilst control males had the same region on the ventral side of the wing blocked.

Males were randomly marked using a black Sharpie marker with an 'x' on either their left or right wing for identification purposes during the experiment. In Panama, experiments began at 8.30am and males were left in the cage until 3pm. During mating, *Heliconius* pairs invariably remain connected for at least an hour and so observations were made every hour to check for matings. If no mating occurred on the first day, this was repeated the next day with the same butterflies. Behavioural observations were

recorded for 17 trials with *H. erato demophoon* and 31 trials with *H. melpomene rosina* for the first two hours of the experiment on day one. Observations were divided into one minute intervals, during which both female and male behaviours were recorded. In Colombia, experiments were conducted from 7am to 1pm, checking every 30 minutes for matings. As before, if no mating occurred on the first day, the experiment was repeated the next day with the same butterflies. Female behavioural observations were recorded for 17 trials with *H. timareta florencia* and 18 trials with *H. melpomene malleti* for the first two hours of the experiment on day one. Observations were divided into one minute intervals and were recorded only when a male was actively courting the female. Four female behaviours were recorded: 'Flutter' refers to a high frequency flutter of the wings with a raised abdominal position carried out when another butterfly is in close proximity, which has typically been interpreted as a rejection behaviour (Klein and Araújo 2010; Jiggins 2017). 'Wings open' refers to when the female is alighted with wings open and abdomen raised but without wing fluttering; 'Abdomen up' refers to when the female is alighted with wings closed and abdomen concealed within the wings; 'Fly away' refers to when the female flies away from the male. Of note, 'Flutter' behaviour was only observed when a male was actively courting the female. Male courtships, previously defined as hovering directly over the female (Klein and Araújo 2010), were recorded. Mating outcome results were analysed with binomial tests. We used generalized linear mixed models (GLMMs) with a binomial error distribution and logit link function to test whether females respond differently to control and experimental males. The response variable was derived from trial minutes in which males courted where females performed a particular behaviour ('success') or did not ('failure'). Significance was determined with likelihood ratio tests comparing models with and without male type included as an explanatory variable. Individual female was included as a random effect in all models to avoid pseudoreplication. All statistical analyses were performed with R version 3.3.1 (R Core Team 2018), along with the packages ggplot2 (Wickham 2009), car (Fox and Weisberg 2011) and binom (Dorai-Raj 2014).

Results

Morphological analysis

We identified a marked sexual dimorphism in scale structure (Fig. 2). In the central region of the male hindwing androconia along vein $Sc+R_1$ we identified specialised scales (Fig. 2A), which were absent in females and in the forewing overlap region of males (Fig. 2B, 2C). These scales had brush-like structures at their distal end (Fig. 2D), and were not detected in any other wing region examined. The brush-like scales were found in alternating rows with scales with a normal structure. Moving away from the $Sc+R_1$ wing vein, the density and width of these scales decreased, with isolated brush-like scales found completely surrounded by normal scales. In addition, the base of these brush-like scales was more swollen and glandular as compared to other scales (Fig. 3).

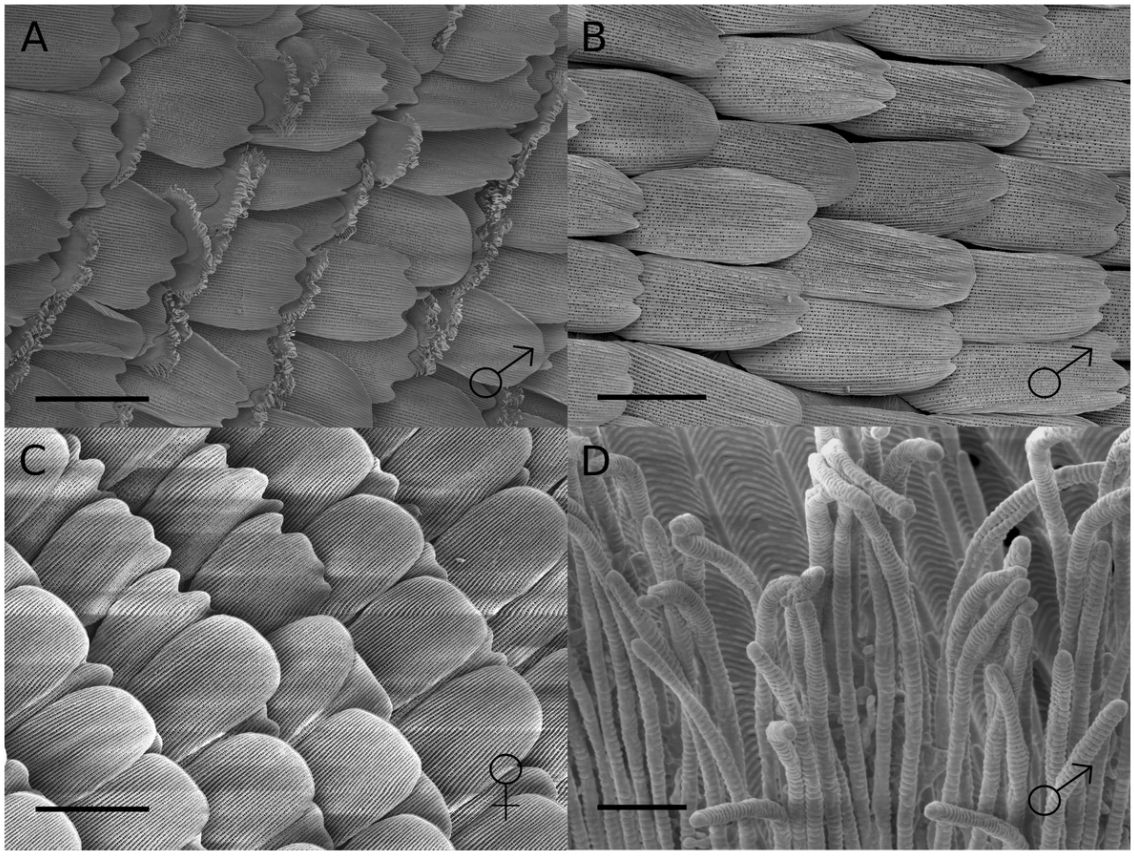


Figure 2: *SEM images of scales from overlap regions of* H. melpomene *wings. (A) Male hindwing; (B) male forewing and; (C) female hindwing at 500× magnification. (D) Magnified view of brush-like structures of the special scales in the male hindwing androconial region. Scale bars indicate 50 µm (A–C) and 2 µm (D).*

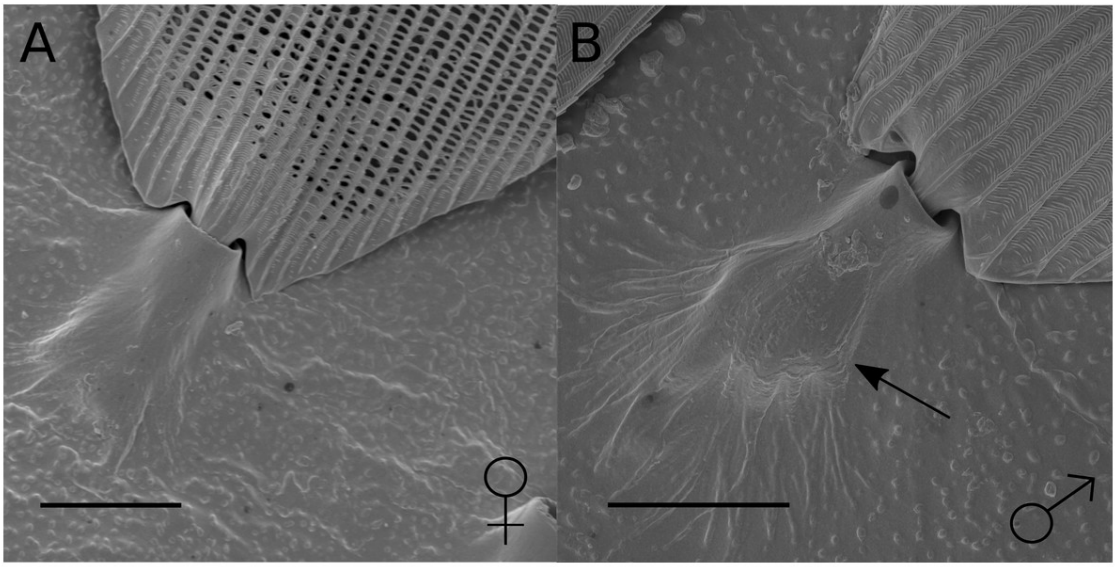


Figure 3: *SEM images of scales from hindwing overlap region in female and male* H. melpomene*. (A) Scale from wing-overlap region of female. (B) Scale from androconial region of male with brush-like structures; the arrow highlights the bulge in the scale base in this region. Scale bars indicate 10 µm.*

Characterization of potential male sex pheromone

We initially investigated candidate wing pheromone composition using a stock of butterflies from Ecuador. By use of GC/MS and synthesis, six compounds were consistently found in the male wing extracts from these samples (Fig. 4) that were identified as the aldehydes (*Z*)-9-octadecenal, octadecanal, (*Z*)-11-icosenal, icosanal, and (*Z*)-13-docosenal and the alkane henicosane (C_{21}).

Firstly, a comparison of different wing regions of 10-day old males was carried out (Fig. 5A). Henicosane was found in all regions of the wing and was not considered in further analysis. The amount of (*Z*)-9-octadecenal was not significantly different between area categories. Octadecanal, (Z)-11-icosenal, icosanal and (*Z*)-13-docosenal showed significant differences between wing areas. Post-hoc testing found that these four compounds were significantly more abundant in the hindwing androconia than the rest of the forewing and hindwing, but not the forewing overlap region (see Table S1 for statistical details). We suggest that the hindwing overlap region be referred to as the androconial region, based on morphological and chemical analyses. The potential role of the, forewing overlap region as an androconia remains to be demonstrated.

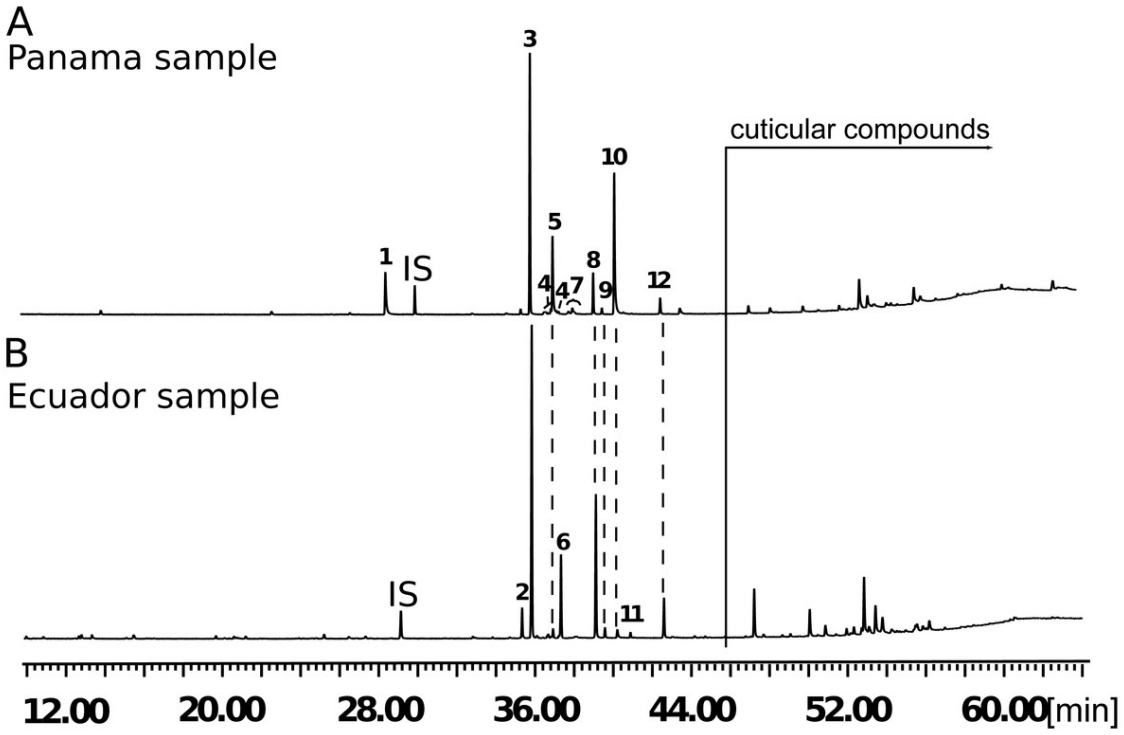


Figure 4: *Regional differences in male androconial extracts. Total ion chromatogram of extract from the androconial region of (A) a Panamanian* H. melpomene rosina *hindwing and (B) an Ecuadorian* H. melpomene *hindwing. 1, syringaldehyde; 2, (Z)-9-octadecenal; 3, octadecanal; 4, methyloctadecanals; 5, 1-octadecanol; 6, henicosane; 7, methyloctadecan-1-ols and nonadecanal; 8, (Z)-11-icosenal; 9, icosanal; 10, (Z)-11- icosenol; 11, tricosane; 12, (Z)-13-docosenal. All peaks eluting later than 44 min are cuticular compounds consisting of larger n-alkanes, 2,5-dialkyltetrahydrofurans, cholesterol or are contaminations. IS, internal standard.*

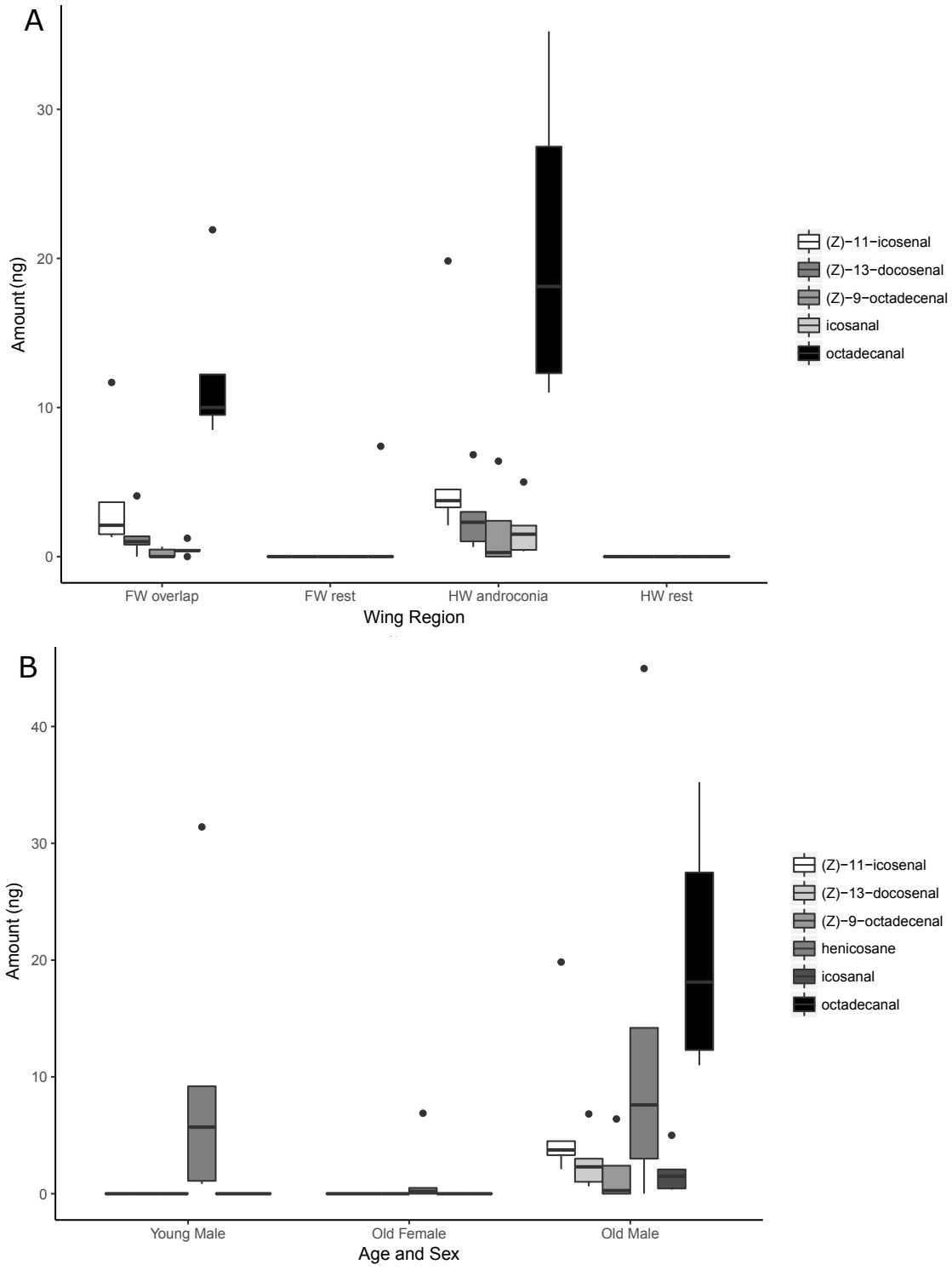


Figure 5: *Compounds detected by GC/MS of* H. melpomene *(Ecuador samples) wing extracts. (A) Presence of compounds in different wing regions of five males (10 days post-eclosion). (B) Presence of compounds in five females (10 days post-eclosion), five young males (0 days post-eclosion) and five old males (10 days post-eclosion).*

Secondly, the hindwing overlap region of old males, old females, and young males were compared (Fig. 5B). With the exception of henicosane, the other compounds were observed to be age-specific and sex-specific. (*Z*)-9-octadecenal was found more in old males than young males or old females but this was not statistically significant. In contrast, octadecanal, (*Z*)-11-icosenal, icosanal and (*Z*)-13-docosenal showed significant differences between age and sex categories. Post-hoc testing found that these compounds were all present in significantly greater amounts in old males than young males or old females (See Table S2 for statistical details).

As stocks of *H. melpomene rosina* from Panama were available for behavioural assays, we then investigated the chemical composition of this population, using a larger sample size. These Panama samples showed some similarities to the Ecuadorean samples, although they contained more compounds and in higher amounts (Fig. 4; Fig. S2 and Table S3). Females and males grouped separately with NMDS visualisation, and these groups were significantly different (Fig. S2). In this larger dataset, (*Z*)-9-octadecenal, octadecanal, (*Z*)-11-icosenal, icosanal, (*Z*)-13-docosenal and henicosane were all found in significantly larger amounts in old males than old females, along with many other compounds (Table S3). Small amounts of nonadecanal, methyl-branched octadecanals and their respective alcohols occurred that had not been detected in the Ecuador samples, potentially due to the difference in equipment sensitivity, genuine geographic variation, or the fact that the Ecuadorean butterflies have spent more generations in captivity. Additionally, syringaldehyde was present, which was not detected in the Ecuador samples.

Behavioural experiments

In our mate choice trials, females of all four species/races discriminated against conspecific males in which pheromone transmission was experimentally blocked (Table 1). Across all four taxa tested, only seven of 71 matings (9.8%) were with the pheromone blocked male, with the remaining 64 matings (90%) being with the control (unblocked) males. This was not due to altered male courtship attempts as control and experimental males courted equally in three out of four species (Fig. S3). In experiments with *H. timareta florencia*, the control males courted more than experimental males (Fig. S3).

When data from *H. timareta florencia* are excluded, 48 out of 52 matings (92%) occurred with the control (unblocked) males.

Table 1: *Outcome of mate choice trials across different species/races. Proportion of successful copulations with the control male was tested using an exact binomial test. Females mated significantly more with the control male than the experimental (pheromone-blocked) male in all four populations. Statistical analysis based on females which mated.*

Species	Mated with control	Mated with experimental	Did not mate	p-value (Exact Binomial Test)
H. melpomene rosina	15	0	18	<0.001
H. erato demophoon	14	1	31	<0.001
H. melpomene malleti	19	3	8	<0.001
H. timareta florencia	16	3	5	<0.01

We observed no consistent significant differences in female behavioural responses towards control and experimental males (Fig. 6). Some female behaviours were observed more often towards experimental males in experiments with *H. melpomene malleti* and *H. timareta florencia* (Fig. 6). In particular, *H. melpomene malleti* females were more likely to open their wings towards experimental males ($2\Delta lnL=17.093$, d.f.=1, $p<0.001$), fly away ($2\Delta lnL= 8.0356$, d.f.=1, $p<0.01$) and also flutter ($2\Delta lnL= 15.823$, d.f.=1, $p<0.001$). Similarly, *H. timareta florencia* females were also more likely to open their wings towards experimental males ($2\Delta lnL= 22.909$, d.f.=1, $p<0.001$), fly away ($2\Delta lnL= 6.1368$, d.f.=1, $p<0.001$), and flutter ($2\Delta lnL= 26.037$, d.f.=1, $p<0.001$). To ensure that behavioural trials without successful mating were not skewing our analysis of female behaviours, we additionally analysed differences between males that mated with the females versus those that did not mate (including those from experiments without matings). Differences between female behaviour towards mated versus unmated males did not differ from differences seen in behaviour towards experimental versus control males (Fig. S4). However, despite these differences, these behavioural responses were not consistently observed across the four species/races tested. Some of these results are driven by just a few individual females, with differences in behaviour no longer significant when they are removed. Furthermore, although some significant female behavioural responses were seen for *H. melpomene malleti* in Colombia, no corresponding difference was found for *H. melpomene rosina* in Panama, where a larger number of courtships were observed, so caution should be taken when interpreting these results.

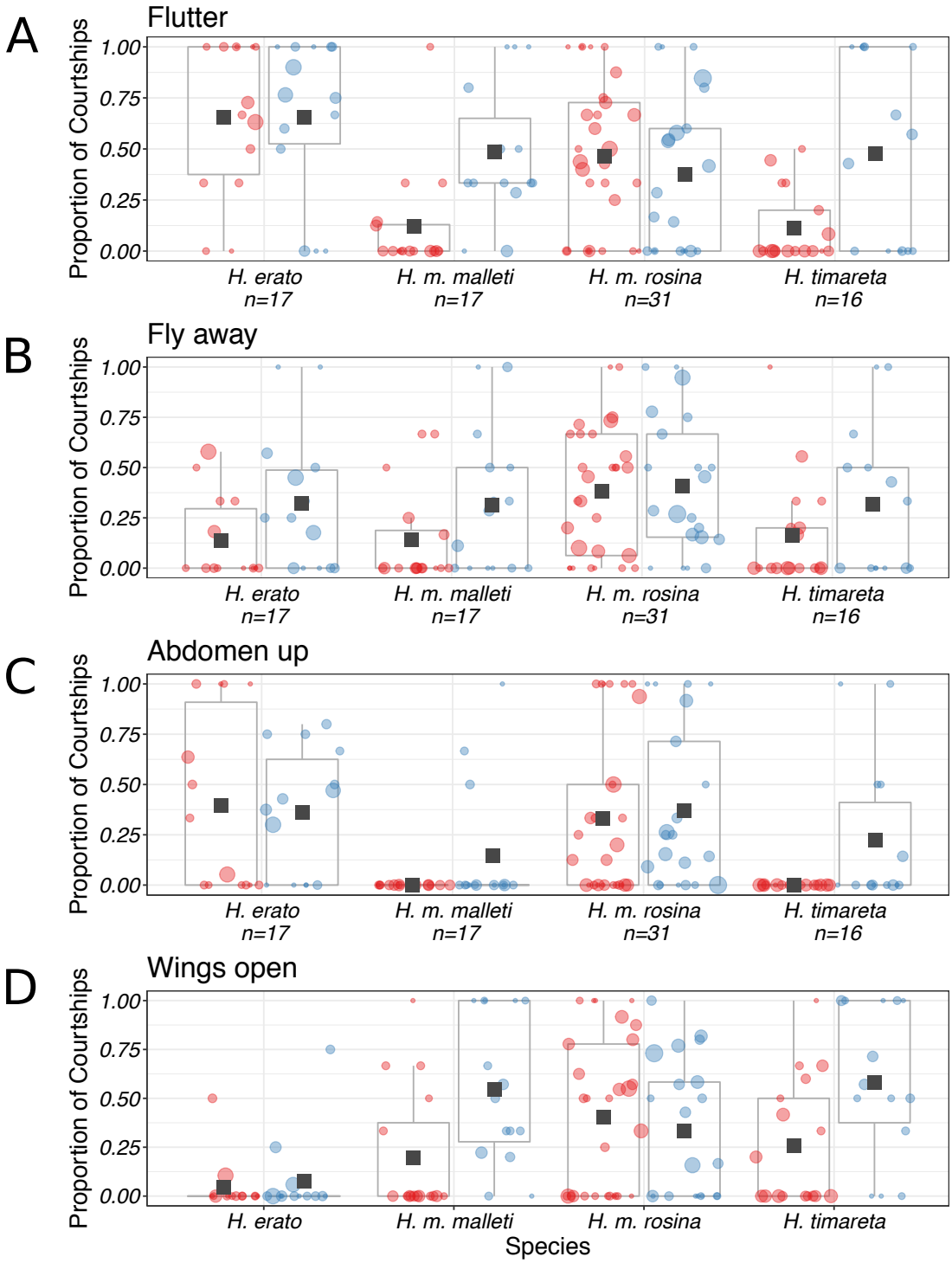


Figure 6: *Proportion of courtships which resulted in different female behavioural responses. Control males are represented in red (left) and experimental males in blue (right). Means are marked with a black square and boxplots mark the inter-quartile ranges. Size of datapoint is proportional to the number of courtships by that male. Female behavioural responses (A) Flutter; (B) Fly away; (C) Abdomen up; (D) Wings open.*

Discussion

Visual cues are known to be important for mate finding and courtship behaviours by male *Heliconius* butterflies, with implications for reproductive isolation and speciation (Merrill et al. 2015). Here, we have shown that female choice based on chemical signalling is also important for reproduction. We have identified compounds associated with sexually mature male wings and described morphological structures putatively involved in pheromone release. Furthermore, we have shown that chemical signalling is involved in mating in *Heliconius*, with females from three different species across four races showing strong discrimination against males which have had their androconia experimentally blocked.

Our results are broadly comparable with another recent analysis of wing compounds in *Heliconius* (Mérot et al. 2015), although the previous study did not compare different wing regions, or males and females of the same age. As the previous study also did not use synthesis to identify compounds, our work is highly complementary and extends their results to confirm region- and age-specific localization of compounds to older male androconia. Male *Heliconius* do not become sexually active until several days after eclosion, so the absence of these compounds from females and younger males is strongly suggestive of a role in mating behaviour. Future experiments will be required to determine if these compounds are sequestered from larval host plants, and if there is genetic control of the production of these compounds, both of which could facilitate a role in reproductive isolation.

The restriction of these five putative male sex pheromones, (Z)-9-octadecenal, octadecanal, (Z)-11-icosenal, icosanal, and (Z)-13-docosenal, to the hindwing androconia of mature males (Fig. 5A) suggests that pheromone storage or production is restricted to the hindwing. This is supported by the scanning electron microscope images which show special brush-like scales in the androconial region (Fig. 2A), located primarily around and along the hindwing vein Sc+R1, similar to the depiction in Figure 73 of Emsley's previous morphological analysis (Emsley 1963). Similar scales have been described from light microscopy in other *Heliconius* species, but not previously in *H. melpomene* (Müller 1912b; Barth 1952). The base of these special brush-like scales was more swollen and

glandular as compared to other scales (Fig. 3), perhaps indicating a role in storage or production of pheromones by these scales. Trace amounts of chemicals on the forewing overlap region may be due to contact in the overlapping portion of the fore- and hindwings, and both wings may play a role in dispersal of the compounds during courtship.

Samples from Panama showed both a greater diversity and amounts of compounds (Fig. S2 and Table S3). This might reflect an issue with inbreeding in the Ecuador population because they were obtained from a commercial breeder, or technical differences between the two locations where the analysis was performed. However, it could also reflect differences in rearing conditions or genuine variation between geographic populations of *H. melpomene*. Further work will be needed to confirm the nature and extent of geographic and individual variation in pheromone composition.

Females exhibited a strong preference for males which did not have their androconia blocked. This suggests that, as in other butterfly systems (Costanzo and Monteiro 2007), female *Heliconius* are actively involved in mating decisions. Nonetheless, there were no consistent differences in the female behaviours we recorded in our experiments. It is possible that the important female preference behaviours are subtle and were missed in our study, or that our sample size may be too small to detect behavioural differences due to individual variation. Alternatively, female acceptance of a male may instead simply represent a decision to stop rejection behaviours, and therefore not be associated with any particular characteristic behavioural response.

It remains unclear which compounds are biologically active and exactly what information is being conveyed. The signal clearly influences female mating decisions, and these compounds may convey complex information about male species identity, quality, age etc. that are interpreted by females. It is also unknown whether females, like males, use visual cues in courtship. The use of multiple signals is common in animal communication (Candolin 2003). Whilst butterflies primarily use visual cues to locate mates (Kemp and Rutowski 2011), it has been shown in *B. anynana* that in addition to visual cues, chemical cues also play a role and are equally important in sexual selection by

female choice (Costanzo and Monteiro 2007). Our work establishes the potential for similar multimodal signalling in *Heliconius* butterflies.

This study provides evidence for the importance of pheromones in intraspecific mate choice in *Heliconius* butterflies. Evolution of cues within populations could lead to reproductive isolation between populations if both cues, and their corresponding preferences, diverge (Ptacek 2000). In other butterflies, male wing compounds contribute to reproductive isolation between closely related species (Grula et al. 1980; Phelan and Baker 1987; Bacquet et al. 2015). Evidence suggests that strong pre-mating barriers in addition to mate preference based on colour wing pattern exist between *Heliconius cydno* and *H. melpomene* (which differ in colour pattern) and between the latter and *H. timareta* (which are mimetic) (Mérot et al. 2017). For example, *H. cydno* males show a preference for their own pattern over that of the closely related *H. melpomene*, but will court wing pattern models of *H. melpomene*. *Heliconius cydno* males, however, have virtually never been observed mating with *H. melpomene* females (Naisbit et al. 2001). On the other hand, although males of the mimetic *H. timareta florencia* and *H. melpomene malleti* equally court female wing models of both species, interspecific matings occur in very low frequency (Sánchez et al. 2015; Mérot et al. 2017). Furthermore, male androconial compounds differ between species (Mérot et al. 2015; Mann et al. 2017).

Supplementary Information

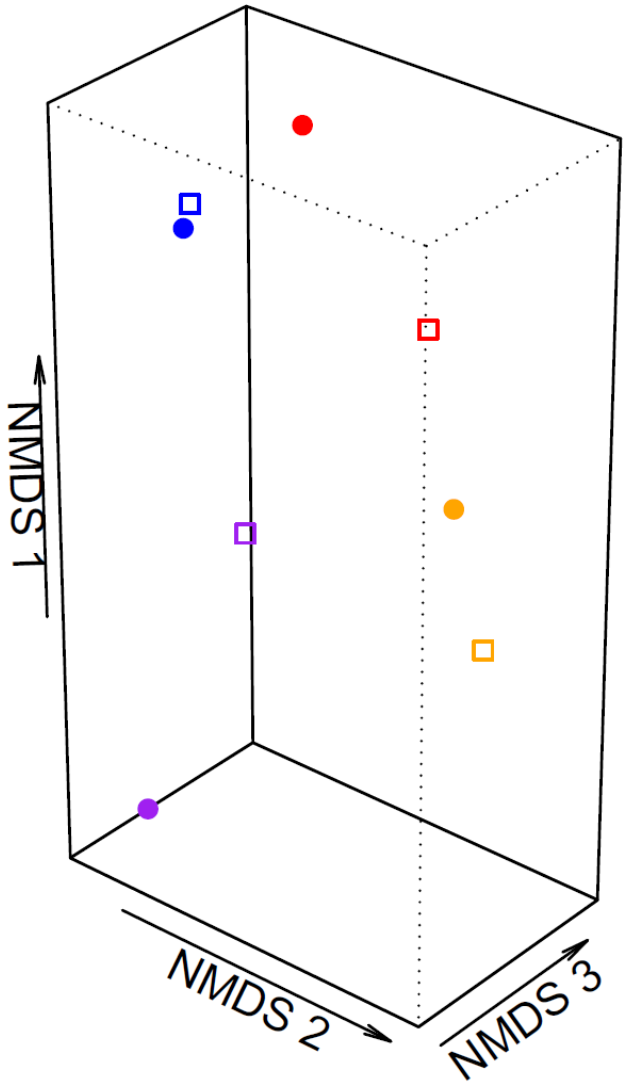


Figure S1: *NMDS in three dimensions on chemical compounds of male hindwings of* H. melpomene *soaked for 1 or 3 hours in solvent (10-12 days post-eclosion, Panama samples). Stress=0.03. Each different colour corresponds to an individual. Closed circles are from the 1-hour group, and open squares 3-hours. No compounds were found in significantly different amounts between groups (Paired t-test, d.f.=4, p=NS).*

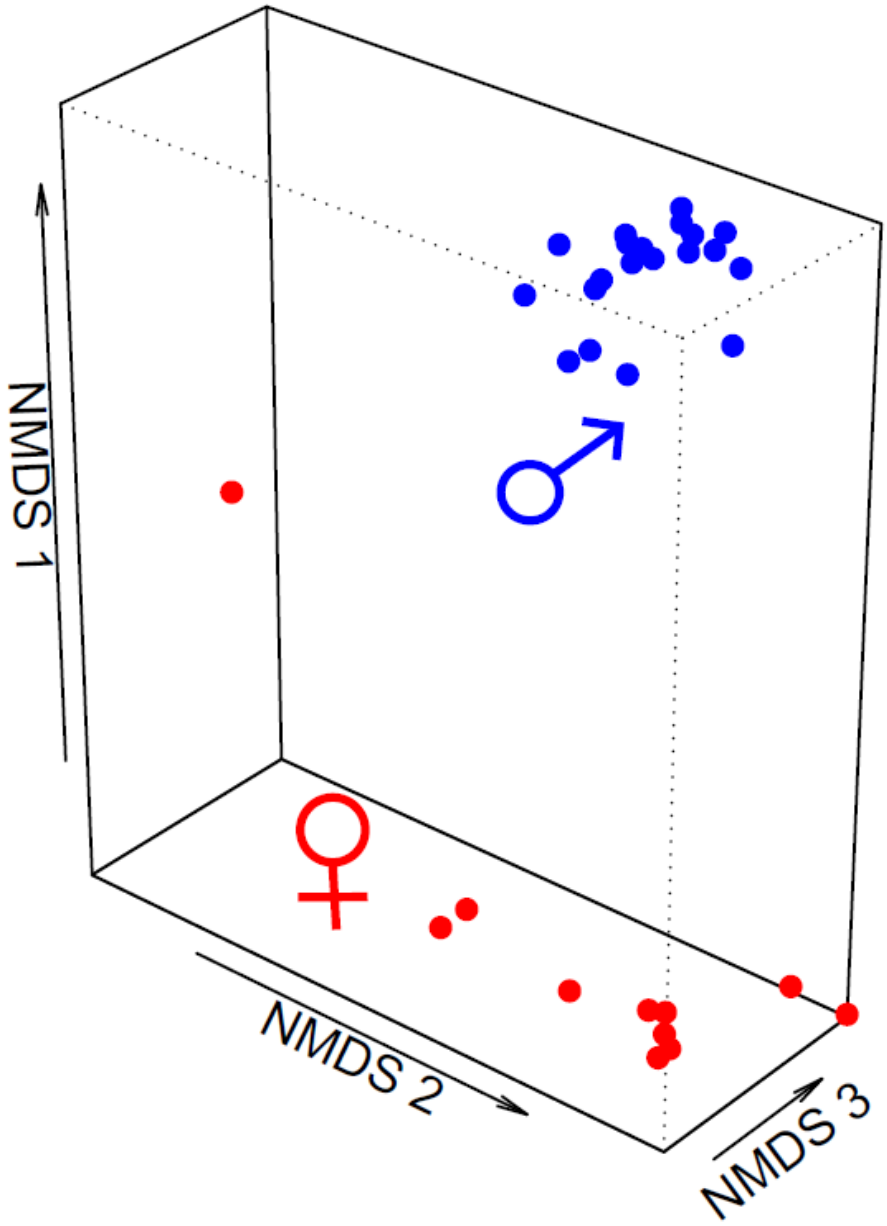


Figure S2: *NMDS in three dimensions on chemical compounds of male and female wings of* H. melpomene *(10-12 days post-eclosion, Panama samples). Stress=0.05. Differences between males and females was supported by ANOSIM (Bray-Curtis R-statistic=0.9226, p=0.001)*

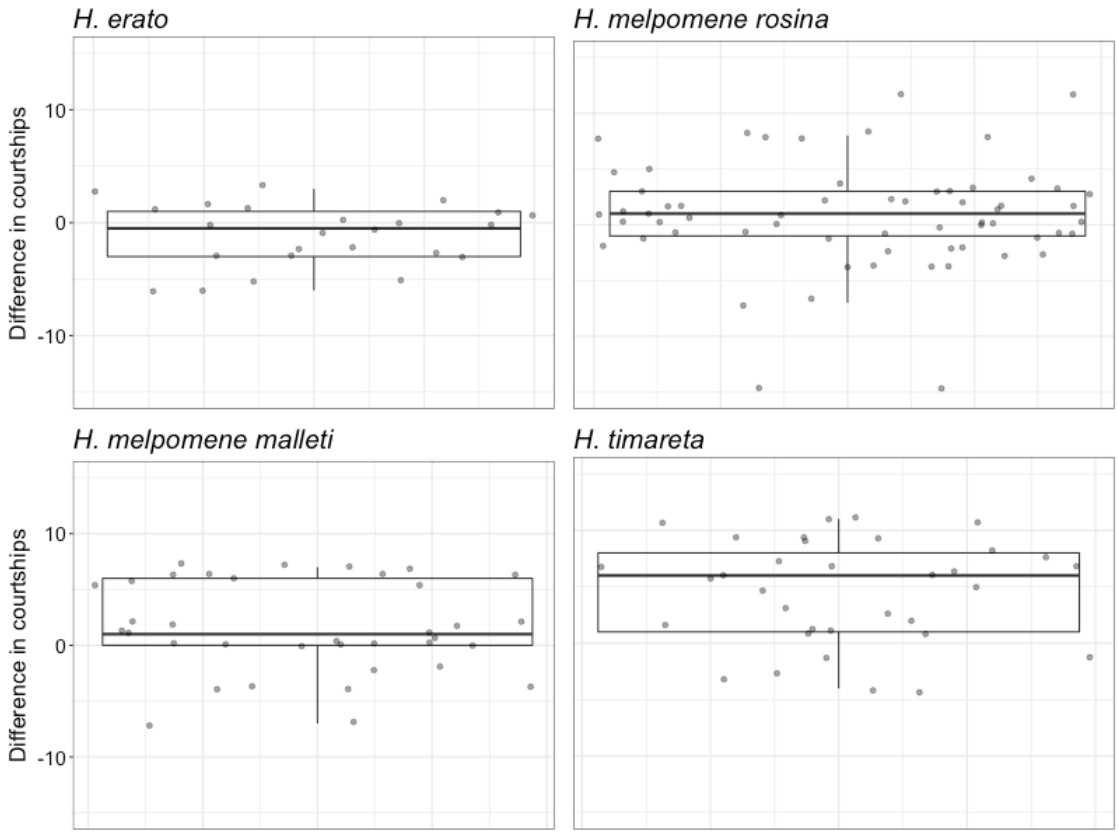


Figure S3: *Difference in number of courtships by control and experimental males represented by boxplots. Each point represents an experiment. Difference not significantly different from zero for* H. erato demophoon*,* H. melpomene rosina *and* H. melpomene malleti*. Control* H. timareta florencia *males courted more than experimental males, complicating the interpretation in this species (Wilcoxin Signed-Rank Test, V=138.5, p<0.01).*

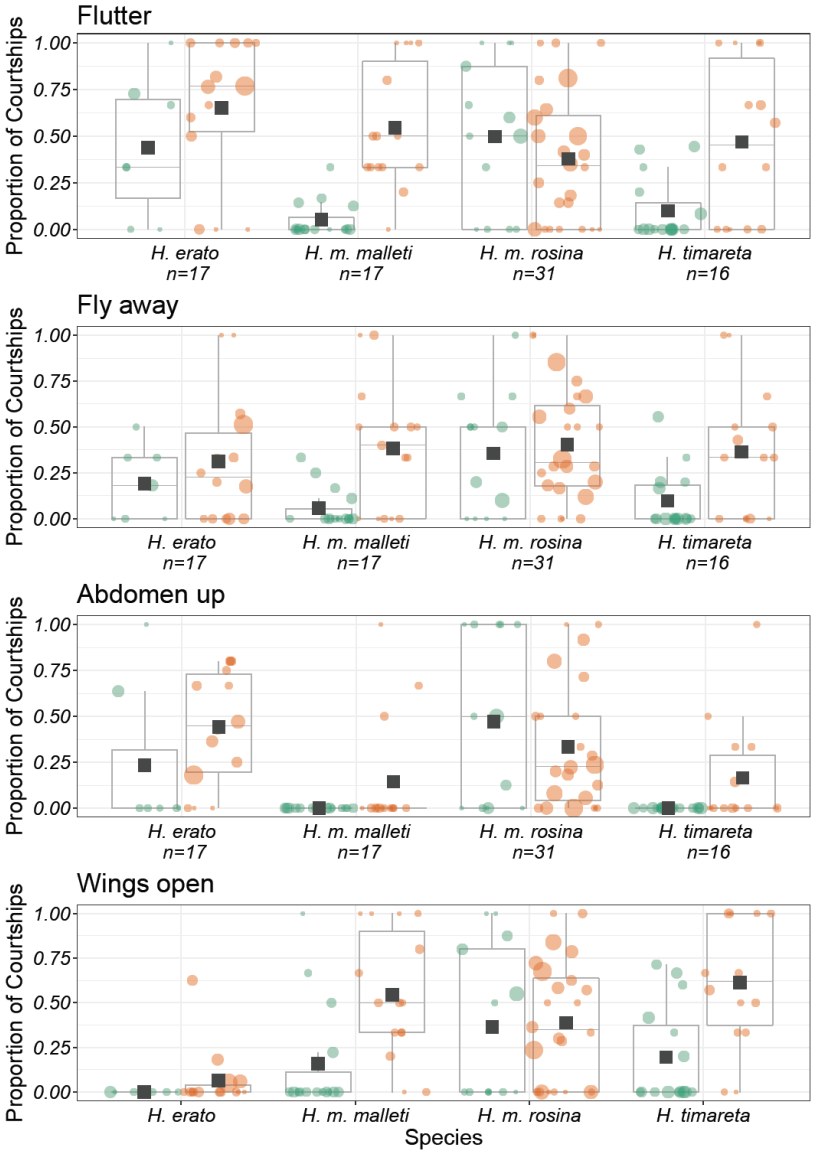


Figure S4: *Proportion of courtships which resulted in different female behavioural responses. Males that mated are represented in green (left) and males that did not in orange (right). Means are marked with a black square and boxplots mark the inter-quartile ranges. Size of datapoint is proportional to the number of courtships by that male. Towards males which did not mate,* H. melpomene malleti *females were more likely to open their (2ΔlnL= 18.65, d.f.=1, p<0.001), fly away (2ΔlnL= 20.454, d.f.=1, p<0.01) and flutter (2ΔlnL= 32.303, d.f.=1, p<0.001).* H. timareta florencia *females were also more likely to open their wings (2ΔlnL= 24.875, d.f.=1, p<0.001), fly away (2ΔlnL= 12.3, d.f.=1, p<0.001), and flutter (2ΔlnL= 23.318, d.f.=1, p<0.001). No significant behavioural differences for* H. melpomene rosina *or* H. erato demophoon *were detected.*

Table S1: *Comparison of amount of each compound in different wing areas of males (10 days post-eclosion, Ecuador samples). FWover refers to the overlapping region of the forewing, FWrest, the remaining portion of the wing, HWand, the hindwing androconial overlapping region, and HWrest, the remaining portion of the wing. Results of Kruskal-Wallis reported for each compound and of post-hoc analysis (Dunn test), where * signifies <0.05, ** <0.01 and *** <0.001.*

				Difference in mean amount of compound between areas					
Compound	x^2	df	p	FWrest-FWover	HWand-FWover	HWrest-FWover	HWand-FWrest	HWrest-FWrest	HWrest-HWand
(*Z*)-9-Octadecenal	6.9	3	NS	-0.228	1.586	-0.228	1.814	-0.000	-1.814
Octadecanal	16.8	3	<0.001	-10.95	8.40	-12.43	19.35*	-1.48	-20.83**
(*Z*)-11-Icosenal	16.8	3	<0.001	-4.05	2.65	-4.05	6.70**	0.00	-6.70**
Icosanal	15.1	3	<0.01	-0.50	1.38	-0.50	1.88**	0.00	-1.88**
(*Z*)-13-Docosenal	14.5	3	<0.01	-1.45	1.31	-1.45	2.76*	0.00	-2.76*

Table S2: *Comparison of amount of each compound in different sex and age categories (mature individuals being 10 days post-eclosion and young 0 days post-eclosion, Ecuador samples). Results of Kruskal-Wallis reported for each compound and of post-hoc analysis (Dunn test), where * signifies <0.05, ** <0.01 and *** <0.001.*

				Difference in mean amount of compound between categories		
Compound	x^2	d.f.	p	Mature female - Young male	Mature male - Young male	Mature male - Young female
(*Z*)-9-Octadecenal	6.9	2	<0.05	0.00	1.81	1.81
Octadecanal	13.3	2	<0.01	0.00	20.83**	20.83**
(*Z*)-11-Icosenal	13.3	2	<0.01	0.00	6.70**	6.70**
Icosanal	13.3	2	<0.01	0.00	1.88**	1.88**
(*Z*)-13-Docosenal	13.3	2	<0.01	0.00	2.76**	2.76**
Henicosane	4.4	2	NS	-8.13	4.30	12.43

Table S3: *Compounds identified in* H. melpomene rosina *(Panama samples) with a mean amount greater than 1 ng in at least one of the sexes. Mean amounts ± standard deviation listed for both male and female butterflies (10-12 days post-eclosion), as well as the percentage of individuals in which the compound was detected. Kruskal-Wallis non-parametric test p-values highlight differences between males and females. Compounds identified by comparison with synthetic samples are shown in bold. Potential contaminants are denoted with an asterisk.*

Chemical	Kovats RI	Male (ng)	%	Female	%	
***o*-Guaiacol**	1090	1.5 ± 1.7	75	0.1 ± 0.2	18	<0.01
Nonanal	1105	24.7 ± 34.2	100	1.5 ± 1.7	73	<0.001
Benzoic acid	1166	2.6 ± 6.7	25	0 ± 0	0	NS
Naphthalene*	1181	1.2 ± 1.1	85	1.0 ± 1.6	36	NS
Decanal	1198	1.7 ± 5.5	30	0.2 ± 0.7	9	NS
Ethyl 4-ethoxybenzoate	1527	7.0 ± 5.1	95	0.7 ± 1.2	27	<0.001
Dihydroactinidiolide	1532	1.1 ± 1.6	60	0 ± 0	0	<0.01
Syringaldehyde	1662	649.9 ± 398.1	100	0 ± 0	0	<0.001
3,5-Dimethoxy-4-hydroxybenzyl alcohol	1707	4.0 ± 6.9	50	0 ± 0	0	<0.01
Unknown hydrocarbon	1709	1.1 ± 1.6	50	0.3 ± 0.6	18	NS
1-(3,5-Dimethoxy-4-hydroxybenzyl)ethanone	1735	1.3 ± 2.0	60	0 ± 0	0	<0.01
Unknown aromatic compound	1738	5.5 ± 5.5	75	0 ± 0	0	<0.001
Unknown aromatic compound	1757	2.1 ± 2.8	55	0 ± 0	0	<0.01
Ethyl benzoate	1762	1.5 ± 1.4	80	1.3 ± 1.3	73	NS
Methyl 1*H*-indol-3-acetate	1822	3.2 ± 3.2	85	0 ± 0	0	<0.001
Unknown compound	1828	0.0 ± 0.1	5	2.1 ± 7.0	9	NS
Methyl 1*H*-indol-3-carboxylate	1853	2.6 ± 2.3	90	0 ± 0	0	<0.001
Unknown compound	1914	1.2 ± 1.8	45	0.2 ± 0.4	27	NS
Unknown aldehyde	1968	6.9 ± 11.0	35	0 ± 0	0	<0.05
Unknown compound	1981	0.0 ± 0.0	0	1.2 ± 3.9	9	NS
(*Z*)-9-Octadecenal	1996	12.1 ± 14.0	80	0 ± 0	0	<0.001
Octadecanal	2021	740.7 ± 411.4	100	0 ± 0	0	<0.001

Unknown compound	2053	2.0 ± 8.8	80	0 ± 0	0	NS
Ethyl 4-hydroxy-3,5-dimethoxybenzoate	2057	41.6 ± 67.5	80	0 ± 0	0	<0.001
Unknown compound	2064	2.9 ± 5.0	30	0 ± 0	0	<0.05
Methyloctadecanal	2065	6.5 ± 6.5	65	0 ± 0	0	<0.001
Henicosadiene	2065	1.3 ± 4.2	10	0 ± 0	0	NS
Methyloctadecanal	2072	4.9 ± 3.7	80	0 ± 0	0	<0.001
Methyloctadecanal	2077	24.4 ± 20.9	65	0 ± 0	0	<0.001
Unknown compound	2078	5.5 ± 24.4	5	0 ± 0	0	NS
Octadecan-1-ol	2082	232.5 ± 149.7	100	0 ± 0	0	<0.001
Henicosane	2099	8.7 ± 16.2	65	0.7 ± 2.5	9	<0.01
Unknown alkene or alcohol	2127	16.3 ± 8.2	90	0 ± 0	0	<0.001
Unknown compound	2132	35.3 ± 22.9	80	0 ± 0	0	<0.001
Unknown compound	2137	6.2 ± 13.1	25	0 ± 0	0	NS
Methyloctadecan-1-ol	2138	7.3 ± 7.3	55	0 ± 0	0	<0.01
Unknown hydrocarbon	2143	4.0 ± 2.4	45	0 ± 0	0	NS
Oleyl acetate	2181	2.0 ± 8.9	5	0 ± 0	0	NS
(*Z*)-11-Icosenal	2198	164.0 ± 127.0	100	0 ± 0	0	<0.001
Icosanal	2224	22.8 ± 11.3	100	0 ± 0	0	<0.001
(*Z*)-11-Icosenol	2261	615.7 ± 304.0	100	0 ± 0	0	<0.001
Tricosene	2281	2.4 ± 2.1	75	0 ± 0	0	<0.001
Unknown compound	2290	1.9 ± 3.5	30	0 ± 0	0	<0.05
Tricosane	2297	1.0 ± 2.4	40	3.2± 4.3	45	NS
Unknown compound	2352	1.9 ± 5.7	15	0 ± 0	0	NS
Unknown compound	2392	11.8 ± 18.3	70	6.8± 7.8	73	NS
(*Z*)-13-Docosenal	2403	64.2 ± 54.6	90	0 ± 0	0	<0.001
Unknown compound	2459	19.9 ± 20.6	65	0 ± 0	0	<0.001
(*Z*)-13-Docosen-1-ol	2464	10.1 ± 23.5	20	0 ± 0	0	NS
Unknown compound	2465	3.3 ± 8.0	20	0 ± 0	0	NS
Unknown compound	2498	0.2 ± 0.7	5	5.8 ± 13.5	18	NS
Pentacosane	2500	3.1 ± 4.7	55	37.6 ± 38.4	82	<0.01
11-Methylpentacosane	2533	0.0 ± 0	0	105.9 ± 111.3	100	<0.001

Chapter 3

Male sex pheromone composition depends on larval but not adult diet in *Heliconius melpomene*

Abstract

Condition-dependent traits can act as honest signals of mate quality, with fitter individuals being able to display preferred phenotypes. Nutrition is known to be an important determinant of individual condition, with diet known to affect many secondary sexual traits. In *Heliconius* butterflies, male chemical signalling plays an important role in female mate choice. Potential male sex pheromone components have been identified previously, although it is unclear what information they convey to the female. In the present study, the effect of diet on androconial and genital compound production is tested in male *Heliconius melpomene rosina*. To manipulate larval diet, larvae are reared on three different *Passiflora* host plants: *Passiflora menispermifolia*, the preferred host plant, *Passiflora vitifolia* and *Passiflora platyloba*. To manipulate adult diet, adult butterflies are reared with and without access to pollen, a key component of their diet. No evidence is found to suggest that adult pollen consumption affects compound production in the first 10 days after eclosion. There is also a strong overlap in the chemical profiles of individuals reared on different larval host plants. The most abundant compounds produced by the butterflies do not differ between host plant groups. However, some compounds found in small amounts differ both qualitatively and quantitatively. Some of these compounds are predicted to be of plant origin and the others synthesised by the butterfly. Further electrophysiological and behavioural experiments will be needed to determine the biological significance of these differences.

Introduction

Sexual ornaments often act as an indicator of mate quality and evolve in response to sexual selection imposed by female preferences (Zahavi 1975, 1977; Andersson 1986). Male "quality" can reflect both direct and indirect benefits gained by the female (Andersson 1994). Direct benefits might include resources that increase female lifetime reproductive success, such as food, shelter, parental care or protection from predators. Indirect benefits, on the other hand, are those that increase genetic quality of a female's offspring. In this case, sexually selected traits reflect the ability of males to provide genes that increase the survivorship or mating success of offspring (Andersson 1986, 1994). These traits may be an honest signal of quality if they are condition-dependent, where only the best quality males are able to display the phenotype.

Male pheromones are a good candidate as an honest signal. Diet-mediated changes can enforce signal reliability (Henneken et al. 2017), and compounds can be costly to produce (Johansson et al. 2005; Harari et al. 2011). Nutritional condition affects male pheromone production in *Tenebrio* beetles (Rantala et al. 2003), cockroaches (Clark et al. 1997) and burying beetles (Chemnitz et al. 2015). Diet manipulation studies show that diet not only affects pheromone production but that these changes affect female mate choice (e.g. cockroaches: South et al., 2011; fruit flies: Liedo et al., 2013).

In addition to overall diet quality, specific diet components can be important. Compounds sequestered in the diet can act directly as sex pheromones, or as pheromone precursors which are then metabolised further (Landolt and Phillips 1997). One well-studied example of this is the sequestration of pyrrolizidine alkaloids (PAs) to make hydroxydanaidal by males of the moth *Utethesia ornatrix* (Conner et al. 1981; Eisner and Meinwald 2003). Some PAs are transferred to the female during mating and chemically protect the eggs, with the male pheromone signalling a direct benefit (Dussourd et al. 1988, 1991; Eisner and Meinwald 1995; Iyengar et al. 2001). In many cases it is probable that both overall nutrient condition and the consumption of specific compounds are both important, such as in the oriental fruit fly, where both overall protein intake and the intake of a specific precursor, methyl eugenol, affect mating success (Shelly et al. 2007).

Diet shifts can provide species with new ecological and evolutionary opportunities. Unique among butterflies, *Heliconius* are able to feed on pollen (Gilbert 1972). They collect pollen from flowers and masticate it on their proboscis to extract amino acids. *Heliconius* have a long lifespan in the wild, facilitated by pollen feeding, which is important for oviposition and viability (Dunlap-Pianka et al. 1977). The lack of dependence upon larval resources for reproduction may have facilitated a greater investment in defensive compounds during the larval stage (Cardoso and Gilbert 2013). As larvae, *Heliconius* caterpillars feed on the cyanogenic leaves of *Passiflora*. *Heliconius* butterflies produce cyanogenic compounds *de novo*, making them unpalatable, and some are also able to sequester compounds directly from *Passiflora* plants (Engler-Chaouat and Gilbert 2006). Both larval and adult diet play important roles in *Heliconius* biology, affecting reproductive lifespan and palatability.

The importance of diet for chemical signalling in *Heliconius* is unclear. The role of chemical signalling in mate choice has been best studied in *Heliconius melpomene rosina* Boisduval (Nymphalidae), a subspecies of *Heliconius melpomene* found in central Panama. Potential male sex pheromone components have been described in the wing overlap region of sexually mature males (Darragh et al. 2017). The morphology of this wing region is sexually dimorphic, with specialised androconial scales only found on male wings. A bouquet of compounds, including octadecanal as a main component, is found in males but not females (Mérot et al. 2015; Darragh et al. 2017; Mann et al. 2017). These chemical cues are important for mating, with females strongly discriminating against males which have their androconia experimentally blocked (Darragh et al. 2017). However, it is still unclear what information (e.g. age or male quality) is being conveyed to females by these cues.

In contrast to androconial compounds which are thought to be aphrodisiac in nature, *Heliconius* males also store anti-aphrodisiac compounds in genital scent glands (Gilbert 1976; Schulz et al. 2007, 2008; Estrada et al. 2011). These are transferred to the female and repel males, delaying re-mating (Gilbert 1976; Schulz et al. 2008). In *H. melpomene,* (*E*)-β-ocimene acts as an anti-aphrodisiac (Schulz et al. 2008). Reduced harassment by other males is thought to be beneficial to the female, and so these compounds could lead to a direct benefit to females. Longer-term, there may be conflict

over the timing of re-mating, as supported by the rapid evolution of genital chemical composition (Andersson et al. 2000, 2004; Estrada et al. 2011). Despite this clear role of genital compounds in male deterrence, the role of these same compounds in female choice remains unknown. Females may benefit from choosing males that have a lower amount of (*E)*-β-ocimene, allowing them to re-mate again sooner. Whilst the dynamics of the costs and benefits of this are unclear, it is quite likely that the genital compounds are involved in female choice.

Both larval and adult diet could be important for the production of androconial and genital compounds as they are not present in freshly-eclosed males (Schulz et al. 2008; Darragh et al. 2017). Feeding experiments with chemically labelled precursors showed that the anti-aphrodisiac compound, (*E*)-β-ocimene, can be synthesized by adult *H. melpomene* via the terpene biosynthetic pathway (Schulz et al. 2008). Pollen intake could be important in providing an energy source for production. Host plant use could affect the chemical bouquet if larval sequestration of specific compounds, or compound precursors, from the host plant is necessary. *Heliconius* raised on their preferred host plant may have a higher quality diet (Smiley, 1978), and so compound production could also be increased due to higher overall quality of the individual.

Here, we investigated how larval and adult diet affect the chemical profile of male *H. melpomene rosina* from central Panama. In Panama, *H. melpomene rosina* females oviposit almost exclusively on *Passiflora menispermifolia* (Merrill et al., 2013). We reared larvae on three different *Passiflora* species: *P. menispermifolia*, the preferred host plant, *P. vitifolia*, and *P. platyloba*. The latter two species are not used by *H. melpomene* in the wild in Panama but are found within the range of *H. melpomene rosina*. These species are therefore potential hosts and larvae survive well on both (Merrill et al., 2013). *H. melpomene* reared on its preferred host plant may have increased energy sources to dedicate to compound production during its adult life. We predict to find both qualitative and quantitative differences in the chemical bouquets of adults reared on different host plants. In a second experiment, we maintained adult male *H. melpomene rosina* with and without access to pollen. We predict that males reared without pollen reduced compound production. In both experiments, we analysed chemical extracts from both the androconial and genital regions of sexually mature male butterflies.

Methods

Butterfly stocks

Heliconius melpomene rosina were reared under ambient conditions at the Smithsonian Tropical Research Institute (STRI) facilities in Gamboa, Panama. Outbred stocks were established from wild individuals collected in Gamboa (9˚7.4' N, 79˚42.2' W, elevation 60 m) in the nearby Soberania National Park and in San Lorenzo National Park (9°17'N, 79°58'W; elevation 130 m). Individuals for this study were reared between February 2016 and April 2017.

Effects of larval diet

Larvae were reared on either *Passiflora platyloba, P. vitifolia* or *P. menispermifolia* (preferred host plant). Adult butterflies were kept in cages with other males and were provided with a ~20% sucrose solution containing bee pollen (Apiarios Malivern, Panama) and with *Psychotria poeppigiana, Gurania eriantha*, *Psiguiria triphylla*, and *Psiguria warscewiczii* as pollen sources. We collected 42 androconial samples from adult butterflies: 19 reared on *P. platyloba*, 11 on *P. menispermifolia*, and 12 on *P. vitifolia.* We also collected 43 genital samples from adult butterflies: 17 reared on *P. platyloba*, 13 on *P. menispermifolia*, and 13 on *P. vitifolia*. We aimed for a minimum of 10 individuals in each group. Variance between groups was due to availability of host plants during the experiment and rearing difficulties on different host plants. Variance between number of androconial and genital samples within each group is due to issues with contamination of samples, which were then unsuitable for analysis. The 19 androconial samples from larvae reared on *P. platyloba* have been previously published (Darragh et al., 2017).

To account for a potential difference in growth rate of individuals reared on different host plants, we measured the forewing length of adult butterflies for the host plant experiments. Before cutting the wings for chemical analysis, we photographed wings beside a ruler. We used ImageJ to calculate forewing length, calibrating the size using the ruler (Schneider et al. 2012).

Effects of adult diet

All larvae were reared on *P. platyloba*, as this was the plant that was available to us in the largest quantities. Adults were then randomly divided into two groups. The first was provided with a ~20% sugar solution containing bee pollen (Apiarios Malivern, Panama) and *P. poeppigiana*, *G. eriantha, P. triphylla*, and *P. warscewiczii* as pollen sources; the second group was only provided with a ~20% sugar solution and no pollen source. We analysed the androconia of 20 individuals reared with pollen and 33 without, and also the genitals of 20 individuals reared with pollen, and 27 without.

Extraction and chemical analysis of tissues

Chemical extractions were carried out on androconial and genital tissue of mature male individuals (10-12 days post-eclosion). The individuals raised on *Passiflora platyloba* for the host plant experiment are previously published samples (Darragh et al., 2017). Genitals were removed using forceps. The wings of the individual were then removed. The hindwing androconial region of the wing, previously described as the grey-brown overlapping region of the wing (Darragh et al., 2017), was dissected from the wings for analysis (Darragh et al. 2017). To extract compounds, the tissue was soaked, immediately after dissection, in 200µl dichloromethane containing 200ng 2-tetradecyl acetate (internal standard) in 2ml glass vials with PTFE-coated caps (Agilent, Santa Clara, USA) for one hour. The solvent was then transferred to new vials and stored at -20°C. Samples were evaporated in the laboratory at room temperature prior to analysis.

Chemical extracts were analysed by GC/MS using an Agilent (model 5977) mass-selective detector connected to an Agilent GC (model 7890B). This was equipped with an Agilent ALS 7693 autosampler and an HP-5MS fused silica capillary column (Agilent, Santa Clara, USA, 30 m × 0.25 mm, 0.25 µm). Injection was performed in splitless mode (250°C injector temperature) with helium as the carrier gas (constant flow of 1.2 ml/min). The temperature programme started at 50°C, was held for 5 minutes, and then rose at a rate of 5°C/minute to 320°C, before being held at 320°C for 5 minutes. Components were identified by comparison of mass spectra and gas chromatographic retention indices with those of authentic reference samples and also by analysis of mass spectra. Components

were quantified using 2-tetradecyl acetate as an internal standard. Only compounds eluting earlier than hexacosane were analysed in androconial samples (Darragh et al., 2017). Later compounds were identified as cuticular hydrocarbons, 2,5-dialkyltetrahydrofurans, cholesterol and artefacts (e.g. phthalates or adipates). The variability in the late eluting cuticular hydrocarbons was low and did not show characteristic differences between samples.

Statistical analysis

We visualised the data using a non-metric multidimensional scaling (NMDS) ordination, based on a Bray-Curtis similarity matrix, in three dimensions. This was carried out using the metaMDS function in the R-package vegan version 2.5-1 (Oksanen et al. 2017), with visualisation using the R-package ade4 (Dray and Dufour 2007).

Following visualisation of the data, we used multivariate statistical techniques to investigate differences between the groups. To identify differences in variance between groups we used the betadisper and permutest functions to test for homogeneity of dispersion. To compare overall chemical composition between groups, we carried out PERMANOVA (permutational multivariate analysis of variance) testing. This was performed using a Bray-Curtis similarity matrix, with 1000 permutations, using the adonis2 function in the package vegan (Oksanen et al., 2017). The "margin" option in adonis2 was used to determine the effect of each term in the model, including all other variables, to avoid sequential effects. The dependent variable in the model was the matrix of compounds, with explanatory variables different in the two experiments. In the larval diet experiments we included wing length and plant species as explanatory variables. In the adult diet experiment we tested the explanatory variable of presence or absence of pollen. We repeated the multivariate analysis using relative rather than absolute amounts of compounds.

We followed up the multivariate statistical analysis with univariate analysis to test for differences in amounts of individual compounds between groups. We tested each compound individually using non-parametric Kruskal-Wallis tests. To correct for multiple-testing, we used the p.adjust function in R, with false detection rate (fdr) correction,

which controls for the proportion of false positives. We repeated the univariate analysis using relative rather than absolute amounts of compounds.

To test for differences in forewing length for individuals reared on different host plants we carried out an ANOVA. For post-hoc analysis, Tukey's HSD test was used to determine which groups were significantly different from each other. All statistical analyses were performed with *R* version 3.3.1 (R Core Team 2018).

Results

Effect of host plant on wing size

Larval host plant affected forewing length (ANOVA, $F_{2,42}$=3.755, p=0.032, Figure 1). Adults which were reared as larvae on *P. menispermifolia* (the preferred host plant of *H. melpomene rosina*) had a mean forewing length of 3.54±0.27cm, those on *P. platyloba* 3.52±0.15cm and on *P. vitifolia* 3.36±0.10cm. However, post-hoc Tukey comparisons did not find any pairwise significant difference between groups.

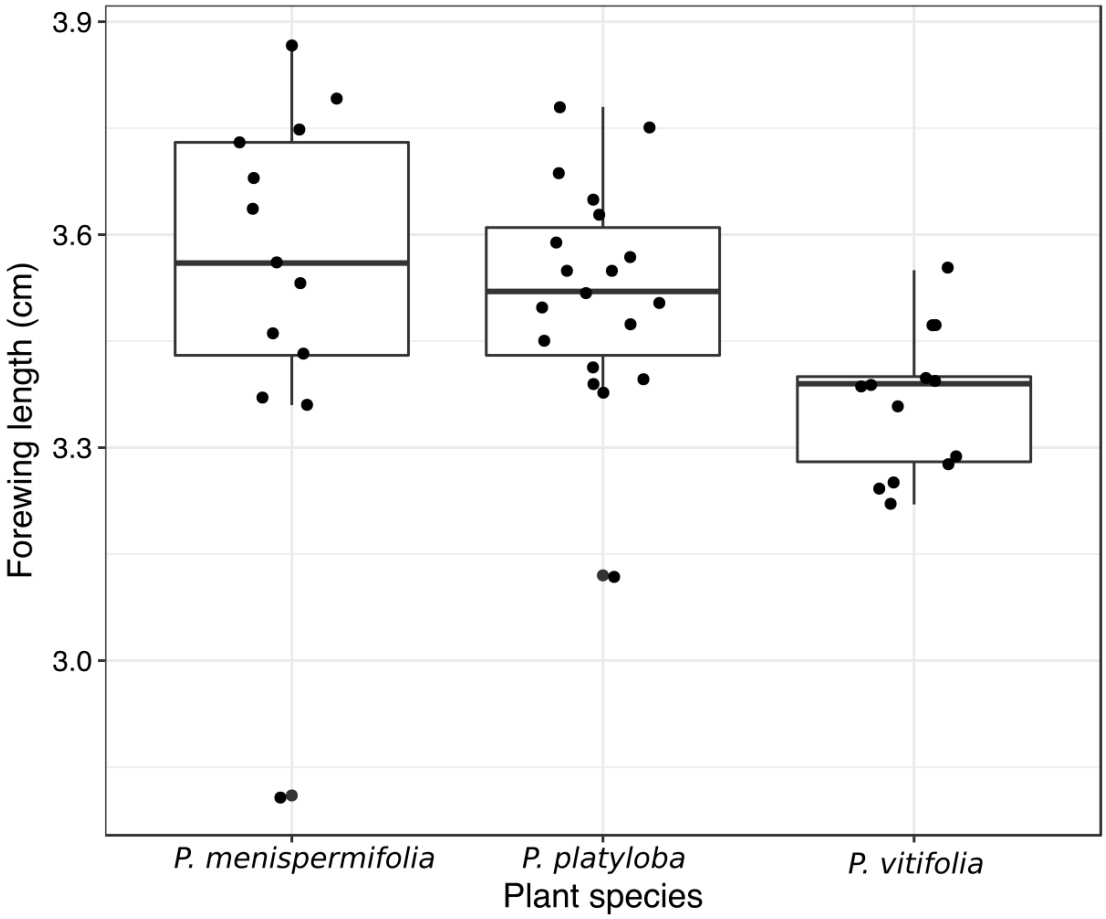


Figure 1: *Larval host plant affects forewing length of* H. melpomene *male adults (ANOVA, df=2, F=3.755, p=0.032). Post-hoc testing using Tukey's HSD found no significant pairwise differences between groups.*

Chemical compounds in androconia and genitals of H. melpomene

We initially analysed the androconia of 19, and genitals of 18, *H. melpomene* reared on *P. platyloba* (Darragh et al. 2017). This is a reanalysis of the 19 androconial samples which have been previously published (Darragh et al. 2017). The most abundant compounds found in the androconia are syringaldehyde, octadecanal, octadecan-1-ol, (*Z*)-11-icosenal, and (Z)-11-icosenol, as found previously (Fig. S1, Table S1)(Darragh et al. 2017). These compounds are present with a mean of greater than 100ng per individual, with octadecanal found in the highest amounts (mean 740ng).

The genital region is dominated by one main compound, (*E*)-β-ocimene, which is found in amounts 20 times greater than any other genital compound (mean 34789ng). This was previously reported in other samples of *H. melpomene* (Schulz et al., 2008). It is found alongside a bouquet of other terpenes, alcohols, aromatic compounds, macrolides, esters and alkanes (Fig. S2, Table S2)(Schulz et al. 2008).

There is little overlap in compounds found between the two body regions, with only ten out of 117 compounds found in both. The genital region contains higher amounts of compounds, and overall more compounds, with 80 in the genitals compared to 47 in the androconia. The most abundant genital compound, (*E*)-β-ocimene, is more volatile (has a higher vapour pressure) than the main compounds found in the androconial region.

Effects of larval diet

Wing size was not a significant factor influencing chemical composition of *H. melpomene* genitals (PERMANOVA, $F_{1,39}$=0.577, p=0.606), but did influence androconial chemical bouquets (PERMANOVA, $F_{1,38}$=3.033, p=0.038), accounting for approximately 7% of the variation.

Our experiments revealed that *H. melpomene* reared on *P. platyloba, P. menispermifolia*, or *P. vitifolia* did not differ significantly in their overall androconial bouquet (PERMANOVA, $F_{2,38}$=1.791, p=0.080, Figure 2A, Table S1), and did not differ in dispersion between groups (permutation test of homogeneity of dispersion, $F_{2,39}$=1.335,

p=0.275). However, when we look at the individual compounds in each treatment, over one quarter (12/47), are present in significantly different amounts between groups (Table 1). These same compounds were also found to be significantly different between groups using relative amounts (for further analysis of relative amounts, see Supplementary Information).

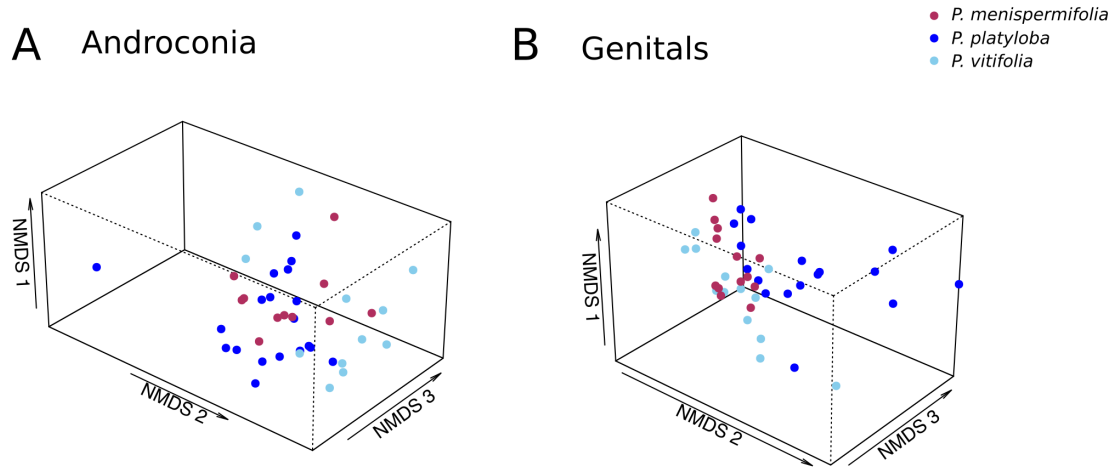


Figure 2. *NMDS (non-metric multidimensional scaling) plot illustrating in three dimensions the overlapping variation in chemical compounds of male* H. melpomene *raised on three different* Passiflora *species.* Passiflora menispermifolia *is the preferred host plant of this species. (A) Androconial compound bouquets do not differ significantly after 10 days. Stress=0.140. (B) Genital compound bouquets do not differ significantly after 10 days. Stress=0.098.*

Table 1. *Androconial compounds that significantly differed between* Heliconius melpomene *reared on different host plants.*

Chemical	RI	*P. platyloba*	*P. menispermifolia*	*P. vitifolia*	H test stat	p-val
Methyl salicylate	1189	0.59±1.97	0.98±3.13	2.32±1.95	10.23	0.026
Unknown compound	1353	0±0	0.40±1.31	3.30±3.85	16.68	0.006
1-(3,5-Dimethoxy-4-hydroxybenzyl) ethanone	1735	0.99±1.58	0.67±0.76	0.01±0.04	10.71	0.022
Benzyl benzoate	1766	0.25±0.57	0.13±0.44	1.27±1.40	10.78	0.022
1-(4-Hydroxy-3,5-dimethoxyphenyl)-2-propen-1-one	1807	1.48±1.93	1.15±1.32	0±0	12.06	0.016
Syringaldehyde derivative	1891	0±0	3.05±2.74	0±0	19.11	0.003
Unknown hydrocarbon	1962	1.32±1.14	1.80±1.41	0.28±0.97	11.19	0.022
Ethyl 4-hydroxy-3,5-dimethoxybenzoate	2057	32.61±55.79	11.60±13.95	0.14±0.37	13.52	0.011
Henicosadiene	2065	1.36±4.28	0.22±0.74	2.97±3.91	9.75	0.030
Methyloctadecanal	2077	30.77±15.99	15.37±8.54	10.94±13.15	13.92	0.011
Icosanal	2224	23.21±11.46	16.08±14.41	8.41±7.59	14.92	0.009
Fatty acid amide	2325	0.27±1.20	0±0	1.09±1.66	12.53	0.015

The gas chromatographic retention index (RI) is reported for each compound. Mean ± SD amounts (ng), as well as Kruskal–Wallis non-parametric test P-values (false detection rate corrected) are provided. Compounds shown in bold are predicted to be plant-derived.

Table 2. *Genital compounds that differed significantly between* Heliconius melpomene *reared on different host plants.*

Chemical	RI	*P. platyloba*	*P. menispermifolia*	*P. vitifolia*	H test stat	p-val
7-β-(*H*)-Silphiperfol-5-ene	1345	9.75±12.23	0±0	0±0	21.46	<0.001
Unknown sesquiterpene	1378	2.87±4.20	0±0	0±0	19.02	0.001
Unknown sesquiterpene	1384	20.90±21.99	0±0	0±0	32.40	<0.001
Unknown compound	1396	39.76±26.86	26.83±30.96	8.23±6.87	16.03	0.004
β-Caryophyllene	1417	18.97±22.09	0±0	0±0	32.40	<0.001
Unknown compound	2250	0±0	1.48±5.32	1.33±2.09	11.37	0.039
Cholestadiene	2744	0±0	9.56±34.48	18.18±24.10	19.96	<0.001

The gas chromatographic retention index (RI) is reported for each compound. Mean ± SD amounts (ng), as well as Kruskal–Wallis non-parametric test P-values (false detection rate corrected) are provided. Compounds shown in bold are predicted to be plant-derived.

Genital compounds of *H. melpomene* reared on *P. platyloba*, *P. menispermifolia*, or *P. vitifolia* did not differ overall between host plant treatments (PERMANOVA, $F_{2,39}$=1.184, p=0.308, Figure 2B, Table S2). The dispersion of individuals between treatments did differ (permutation test of homogeneity of dispersion, $F_{2,40}$=3.668, p=0.034), with pairwise permutation analysis revealing that the dispersion of individuals raised on *P. menispermifolia* is different from both *P. vitifolia* and *P. platyloba*, which do not differ from each other (Table S3). NMDS visualisation reveals less variation between individuals reared on the preferred host *P. menispermifolia* (Figure 2B). Furthermore, seven out of 80 individual compounds differ between groups (Table 2). These compounds were also found to be significantly different between groups using relative amounts, excluding an unknown compound (gas chromatographic retention index, RI=1396) which is no longer significant (for further analysis of relative amounts, see Suppementary Information).

None of the individual compounds found to differ between groups were the most abundant compounds. These compounds are not found in an average amount higher than 40ng per individual, much smaller amount than the most abundant compound in androconia (740 ng) or genitals (34789 ng). We find both qualitative (presence or absence of compound) and quantitative (difference in amount of compound) differences between butterflies reared on different host plants. For example, in the androconial samples, a syringaldehyde derivative is only found in individuals reared on *P. menispermifolia*. Icosanal, in contrast, is found in butterflies reared on all three host plant species but in significantly different amounts (Table 1).

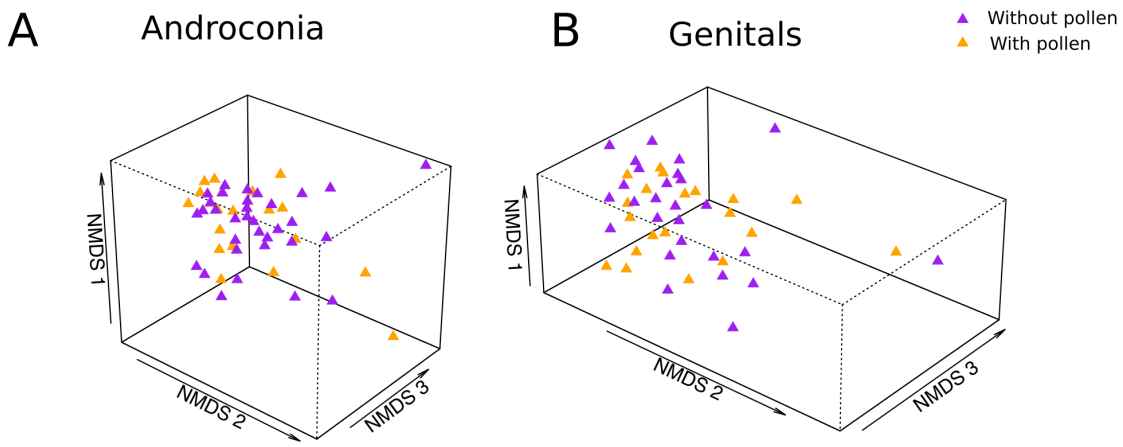


Figure 3. *NMDS (non-metric multidimensional scaling) plot illustrating in three dimensions the overlapping variation in chemical compounds of male* H. melpomene *raised with or without pollen. (A) Androconial compound bouquets do not differ significantly after 10 days. Stress=0.131. (B) Genital compound bouquets do not differ significantly after 10 days. Stress=0.108*

Effects of adult diet

H. melpomene butterflies reared with or without pollen for 10 days do not differ in either androconial (PERMANOVA, $F_{1,51}$=1.653, p=0.145), or genital (PERMANOVA, $F_{1,45}$=1.259, p=0.260) chemical bouquets (Figure 3, Table S4, Table S5). False detection rate corrected Kruskal-Wallis testing found no compounds in significantly different amounts between the groups.

Discussion

Chemical signalling is known to be important for intraspecific female mate choice in *Heliconius* (Darragh et al., 2017). The information conveyed by these compounds, such as age, species identity, or mate quality, remains unclear. We find no evidence that adult pollen consumption affects compound production in the first ten days of adult life. In contrast, individual androconial and genital compounds were found in different amounts between larval host plant treatment groups, and dispersion varied between host plant treatments for genitals.

The most abundant compounds identified in *H. melpomene* androconia and genitals are the same as those identified by previously published studies (Schulz et al., 2008; Estrada et al., 2011; Mérot et al., 2015; Darragh et al., 2017; Mann et al., 2017). We did not identify all the compounds found previously in genitals (Schulz et al., 2008). This is likely due to variation in sample collection, as the previous compound list was derived from pooled samples, allowing for a higher detection threshold. In the androconia, we did not detect ethyl palmitate, ethyl oleate or ethyl stearate, previously reported compounds (Mann et al., 2017). We did detect ethyl oleate in the genitals suggesting that previous reports in the androconia may be due to contamination from genital contact. We also found many more compounds, probably due to improved GC-MS detection thresholds which are more sensitive to compounds found only in low levels.

We did not find a difference in the ability of males to produce compounds when reared with and without pollen in this experiment. This finding was somewhat unexpected, as pollen is an important resource for adult *Heliconius*. This result could

mean that chemical signalling is not nutritionally dependent. *Heliconius* are one of the most long-lived butterflies, with adults known to live more than 8 months in the wild (Gilbert 1972) and pollen limitation might play a more important role over longer time scales. In females it has been shown that the effects of pollen for oviposition and viability are evident after about a month (Dunlap-Pianka et al. 1977), suggesting that until that point, larval reserves are sufficient. This could be the same for males and we therefore cannot rule out the possibility that adult nutrition might influence pheromone production only later in life.

Pollen-feeding is thought to be important for another aspect of reproduction in *Heliconius*: the donation of a spermatophore to the female during mating. The spermatophore is protein-rich, and so the amino acids obtained by pollen-feeding are probably needed to make new spermatophores after each mating (Cardoso and Silva 2015), as supported by the fact that males with more lifetime matings collect more pollen overall (Boggs 1990). This is beneficial for the females, as its reduces the females' need to forage for pollen (Boggs and Gilbert 1979; Boggs 1981, 1990). It has been proposed that females may determine spermatophore quality using cues, such as chemicals, that indicate direct benefits for the female (Cardoso and Silva 2015). Alternatively, females could benefit indirectly through a "good genes" mechanism (Andersson and Simmons 2006), for example through inheritance of foraging ability (Karino et al. 2005). Further experiments will be required to determine if chemical signalling in older males can indicate spermatophore production and male quality.

Chemical profiles produced by adult butterflies reared on the three larval host plants are largely overlapping. *Heliconius melpomene* is able to produce the majority of compounds found in both androconial and genital bouquets when reared on all three *Passiflora* species. The most abundant compounds are not found in significantly different amounts. However, we find less variation in genital compounds produced by individuals reared on *P. menispermifolia*, the preferred host plant of *H. melpomene rosina,* compared to *P. vitifolia* and *P. platyloba*, perhaps suggesting some level of chemical or digestive specialisation.

Despite overall similarity between butterflies reared on different host plants, there are differences in some specific androconial and genital compounds. These differences are both qualitative and quantitative in nature. Over one quarter of androconial compounds are found in significantly different amounts between the three groups, along with almost one tenth of genital compounds. One third of these significant androconial compounds are thought to come from the phenylpropanoid pathway, active in plants (Boerjan et al. 2003). This pathway forms aromatic compounds with an alkyl sidechain of three carbons that serve as building blocks for lignin and lignans. Oxidative degradation of lignine or by-products of this biosynthetic pathway are the source of the aromatic compounds syringaldehyde, 1-(3,5-dimethoxy-4-hydroxybenzyl)ethanone or ethyl 4-hydroxy-3,5-dimethoxybenzoate. They are not currently known to act as pheromones in insects, although closely related compounds, lacking one methoxy group, are reported as fruit fly and moth pheromones (Francke and Schulz 2010).

Over half of the genital compounds, specifically sesquiterpenes, are also thought to originate from plant sources. These include the specific compounds 7-β-(*H*)-silphiperfol-5-ene and β-caryophyllene as well as some unknown other sesquiterpenes. The genome currently does not show any indication of a required sesquiterpene cyclase in *H. melpomene*, thus making de novo synthesis of sesquiterpenes by the butterflies unlikely. These data suggest that differences in plant biochemistry affect the chemicals released from both androconial and genital regions of the adult butterfly.

We do not know which components of the androconial bouquet are biologically important in *Heliconius melpomene*. It cannot be assumed that the most abundant compounds are necessarily the most important, as minor compounds can often play important roles in attraction (D'Alessandro et al. 2009; McCormick et al. 2014) . Furthermore, the response to pheromonal cues is blend-specific in other Lepidopteran systems (Yildizhan et al. 2009; Larsdotter-Mellström et al. 2016). It is therefore possible that, despite the overlap in overall chemical composition between host plant treatments, the compounds which are significantly different could drive a change in female response. This seems particularly likely for the androconial bouquet, where more than a quarter of compounds are found in different amounts between larval host plant treatments. However, it is important to note that direct tissue extraction of chemicals may not

accurately reflect chemical amounts emitted by live butterflies (Visser et al. 2018). Electrophysiological and behavioural experiments will be required to determine if these differences are biologically relevant.

Our results contribute to an understanding of why some populations of *H. melpomene* are host specialists when their larvae can successfully feed on a wider variety of host plants. *Heliconius melpomene* larval growth rate is similar on different host plants under laboratory conditions (Smiley, 1978), but there are slight differences in survival in the wild, perhaps due to ant attendance or parasitism (Merrill et al., 2013). In particular *H. melpomene* fed on *P. vitifolia* show a somewhat lower survival rate when compared with the natural host plant *P. menispermifolia* (Merrill et al., 2013), which may be related with our finding that host plant affects size in *Heliconius melpomene*. If host specialisation is not due to physiological adaptation in the larvae, an alternative is that it could be explained by sexual selection. Female choice on diet-derived pheromones could potentially drive host plant specialisation (Quental et al. 2007). In this case we would not expect larval host plant specialisation as selection is acting on the adult stage. To determine if this could be a plausible mechanism to explain host plant preference in *Heliconius* we would need to determine whether the compounds that change with host plant use are also important for female choice.

We might expect to find intraspecific differences in chemistry between *Heliconius* races. Across their geographic range, *Heliconius* butterflies use different host plants and can vary in their extent of host plant specialisation (Benson et al. 1975; Benson 1978). Based on our results, we predict this to lead to differences in chemical bouquets between populations due to differing host plant use. We might also expect to find more intraspecific differences in chemistry in populations which are more generalist. Future investigations using different geographic races will help us understand the role of diet in the chemistry of *Heliconius* butterflies, and its link to host plant specialisation.

Supplementary Information

Results. Repeat of analyses on relative compound amounts in samples. Results found were similar to analysis of absolute compound amounts.

We repeated the analysis on the relative rather than absolute compound amounts. We found that *H. melpomene* reared on *P. platyloba, P. menispermifolia*, or *P. vitifolia* did differ significantly in their overall androconial bouquet (PERMANOVA, $F_{2,38}$=2.015, p=0.048). This was not detected using absolute amounts, however, it only accounts for 9% of variation. We did not detect a difference in dispersion between groups (permutation test of homogeneity of dispersion, $F_{2,39}$=0.559, p=0.559). All the same compounds found in significantly different amounts were the same as those found using absolute amounts (Table 1).

We found that *H. melpomene* reared on *P. platyloba, P. menispermifolia*, or *P. vitifolia* did not differ significantly in their overall genital compound bouquet (PERMANOVA, $F_{2,39}$=0.592, p=0.797). In contrast to the absolute amount analysis, we did not detect a difference in dispersion between groups (permutation test of homogeneity of dispersion, $F_{2,40}$=1.254, p=0.296). The same compounds found in significantly different amounts were the same as those found using absolute amounts (Table 2), apart from the unknown compound (RI=1396) which is no longer significant.

As found with analysis of absolute amounts, *H. melpomene* butterflies reared with or without pollen for 10 days do not differ in either androconial (PERMANOVA, $F_{1,51}$=2.235, p=0.063), or genital (PERMANOVA, $F_{1,45}$=0.795, p=0.500) bouquets.

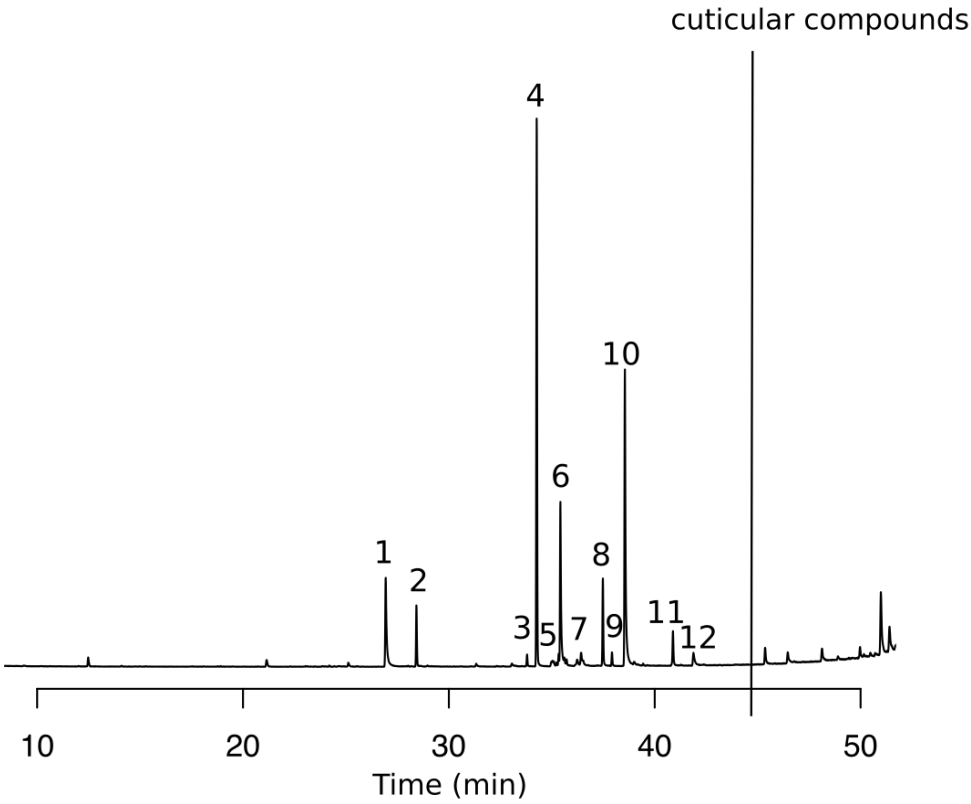


Figure S1. *Total ion chromatogram of extract from androconial region of* H. melpomene*. 1, Syringaldehyde; 2, internal standard; 3, (Z)-9-octadecenal; 4, octadecanal; 5, methyloctadecanals; 6, 1-octadecanol; 7, methyloctadecan-1-ol; 8, (Z)-11-icosenal; 9, icosanal; 10, (Z)-11-icosenol; 11, (Z)-13-docosenal; 12, (Z)-13-docosen-1-ol. All peaks later than 45 minutes are cuticular compounds. Abundances are scaled to the highest peak.*

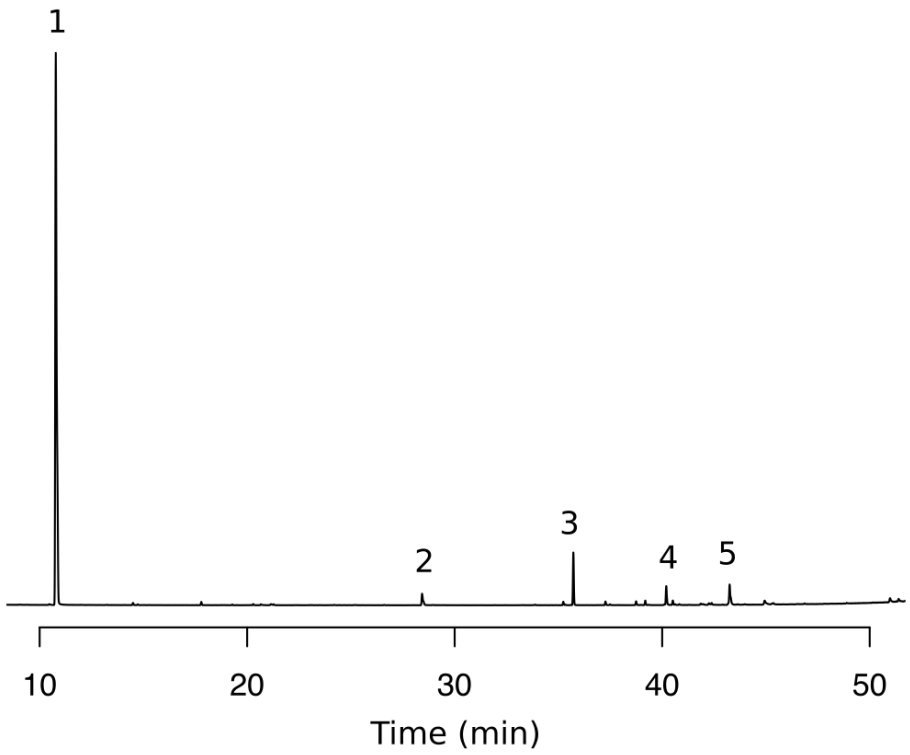


Figure S2. *Total ion chromatogram of extract from genital region of* H. melpomene*. 1,* (E)-*β-ocimene,; 2, internal standard; 3, henicosane; 4, (Z)-3-hexenyl hexadecanoate; 5, hexyl octadecenoate & (Z)-3-hexenyl octadecenoate. Abundances are scaled to the highest peak.*

Table S1: *Androconial compounds identified in* H. melpomene rosina *males reared on* P. platyloba *(19 individuals),* P. menispermifolia *(11 individuals) or* P. vitifolia *(12 individuals), with a mean amount greater than 1 ng in at least one of the treatments. The gas chromatographic retention index (RI) is reported for each compound. Mean amounts (ng) ± standard deviation, as well as false detection rate corrected p-values. Compounds in bold are predicted to be plant-derived.*

Chemical	**RI**	***P. platyloba***	***P. menispermifolia***	***P. vitifolia***	
o-Guaiacol	1090	1.50±1.73	1.52±1.41	1.43±1.28	NS
Nonanal	1105	25.31±34.96	11.64±8.88	10.10±3.91	NS
Benzoic acid	1166	3.61±7.53	0.07±0.16	1.15±1.97	NS
Naphthalene	1181	1.22±1.13	2.17±1.05	1.06±0.65	NS
Methyl salicylate	1189	0.59±1.97	0.98±3.13	2.32±1.95	0.026
Decanal	1198	1.80±5.65	0.86±1.69	0.10±0.34	NS
Unknown compound	1353	0±0	0.40±1.31	3.30±3.85	0.006
Dihydroactinidiolide	1532	1.41±1.65	0.52±0.72	0.48±0.63	NS
Syringaldehyde	1662	641.31±407.13	406.92±228.15	442.11±229.66	NS
3,5-Dimethoxy-4-hydroxybenzyl alcohol	1707	3.67±6.85	1.34±4.45	1.72±2.24	NS
Unknown compound	1709	0.25±0.68	1.16±1.48	0.16±0.35	NS
1-(3,5-Dimethoxy-4-hydroxybenzyl) ethanone	1735	0.99±1.58	0.67±0.76	0.01±0.04	0.022
Unknown hydrocarbon	1753	1.53±2.08	3.46±2.87	1.01±1.72	NS
Ethyl benzoate	1762	2.02±1.99	1.37±1.14	1.52±1.03	NS
Benzyl benzoate	1766	0.25±0.57	0.13±0.44	1.27±1.40	0.022
1-(4-Hydroxy-3,5-dimethoxyphenyl)-2-propen-1-one	1807	1.48±1.93	1.15±1.32	0±0	0.016
Methyl 1*H*-indol-3-acetate	1822	3.37±3.20	3.28±2.79	2.10±2.63	NS
Methyl 1*H*-indol-3-carboxylate	1853	2.57±2.37	1.67±2.54	0.98±1.14	NS
Benzyl salicylate	1870	1.44±2.57	0.57±0.97	1.99±3.67	NS
Syringaldehyde derivative	1891	0±0	3.05±2.74	0±0	0.003

Nonadecane	1898	1.47±2.69	2.48±4.31	0.56±1.04	NS
Unknown compound	1914	1.27±1.80	1.33±1.11	0.78±0.86	NS
Unknown hydrocarbon	1962	1.32±1.14	1.80±1.41	0.28±0.97	0.022
(*Z*)-9-Octadecenal	1996	11.68±14.04	6.34±4.97	7.53±9.54	NS
Octadecanal	2021	739.57±422.66	617.88±536.70	482.05±390.61	NS
Ethyl 4-hydroxy-3,5-dimethoxybenzoate	2057	32.61±55.79	11.60±13.95	0.14±0.37	0.011
Methyloctadecanal	2065	8.11±5.99	5.69±4.96	2.56±3.81	NS
Henicosadiene	2065	1.36±4.28	0.22±0.74	2.97±3.91	0.030
Methyloctadecanal	2072	5.35±3.49	3.65±2.23	2.12±1.90	NS
Methyloctadecanal	2077	30.77±15.99	15.37±8.54	10.94±13.15	0.011
Octadecan-1-ol	2082	231.72±153.72	660.33±861.35	224.31±240.45	NS
(Z)-16-Methyl-9-octadecenol	2092	4.97±4.17	5.67±4.76	2.80±3.43	NS
Henicosane	2099	11.29±15.86	15.19±21.48	9.49±8.28	NS
Unknown alkene or alcohol	2127	16.38±7.55	16.65±13.98	10.54±6.22	NS
Unknown compound	2132	36.70±19.06	45.70±35.74	24.82±14.49	NS
Methyloctadecan-1-ol	2138	9.30±6.07	6.09±8.66	6.25±5.97	NS
(*Z*)-11-Icosenal	2198	168.51±128.84	62.40±46.48	70.95±57.40	NS
Icosanal	2224	23.21±11.46	16.08±14.41	8.41±7.59	0.009
(*Z*)-11-Icosenol	2261	619.09±311.94	514.62±436.15	465.59±229.14	NS
Tricosene	2281	3.98±3.12	8.03±8.92	2.73±2.85	NS
Tricosane	2297	1.19±2.43	1.16±1.30	0.69±1.02	NS
Fatty acid amide	2325	0.27±1.20	0±0	1.09±1.66	0.015
Unknown amide	2374	2.17±6.01	0±0	2.04±4.82	NS
Unknown compound	2392	12.30±18.70	5.30±10.87	1.23±2.86	NS
(*Z*)-13-Docosenal	2403	66.30±54.95	32.82±40.35	32.03±25.53	NS
(*Z*)-13-Docosen-1-ol	2464	31.80±22.13	53.03±93.01	35.28±30.98	NS
Pentacosane	2500	3.43±4.69	1.86±1.84	2.42±1.77	NS

Table S2: *Genital compounds identified in* H. melpomene rosina *males reared on* P. platyloba *(17 individuals),* P. menispermifolia *(13 individuals) or* P. vitifolia *(13*

individuals), with a mean amount greater than 1 ng in at least one of the treatments. The gas chromatographic retention index (RI) is reported for each compound. Mean amounts (ng) ± standard deviation, as well as false detection rate corrected p-values. Compounds in bold are predicted to be plant-derived.

Chemical	**RI**	***P. platyloba***	***P. menispermifolia***	***P. vitifolia***	
β-Myrcene	990	11.94±13.62	15.16±6.25	19.34±31.48	NS
(*Z*)-β-Ocimene	1030	2822.73±11451.08	70.18±27.71	246.54±635.68	NS
(*E*)-β-Ocimene	1054	34789.16±22151.25	35742.69±9716.58	39065.11±30676.99	NS
o-Guaiacol	1090	0.33±0.67	0.34±0.75	1.08±1.81	NS
Alloocimene	1129	8.64±11.64	14.34±14.11	26.66±30.00	NS
2-sec-Butyl-3-methoxypyrazine	1173	41.38±21.95	51.89±27.36	46.52±27.05	NS
2-Methoxy-3-isobutylpyrazine	1181	15.45±9.53	19.99±12.12	17.55±8.85	NS
Methyl salicylate	1189	0.26±1.05	0.63±2.14	1.15±2.22	NS
Dihydroedulan II	1290	35.91±29.64	14.78±11.83	23.38±25.46	NS
7-β-(*H*)-Silphiperfol-5-ene	1345	9.75±12.23	0±0	0±0	<0.001
8-Hydroxylinalool	1349	0±0	0±0	4.07±7.10	NS
Unknown sesquiterpene	1378	2.87±4.20	0±0	0±0	0.001
Unknown sesquiterpene	1384	20.90±21.99	0±0	0±0	<0.001
Unknown compound	1396	39.76±27.38	26.83±30.96	8.23±6.87	0.004
β–Caryophyllene	1417	18.97±22.09	0±0	0±0	<0.001
11-Dodecanolide	1465	2.86±5.50	4.34±7.34	3.84±9.72	NS
14-Tetradecanolide	1733	9.64±19.18	0±0	0±0	NS
Ethyl benzoate	1762	2.18±2.14	1.56±2.63	1.50±3.99	NS
Hexadecenolide	1850	0.27±1.13	0±0	1.36±3.37	NS

15-Hexadecanolide	1855	0.18±0.54	2.87±5.59	0.82±2.09	NS
Hexadecenolide	1865	4.29±13.80	1.28±4.61	1.53±5.52	NS
(*Z*)-9-Octadecenal	1996	0±0	0±0	5.35±1.13	NS
Icosane	1998	10.14±12.67	12.07±6.51	10.97±11.86	NS
Octadecadienolide	2033	0.36±1.01	2.37±5.95	4.27±12.07	NS
(Z)-9-Octadecen-11-olide	2038	5.88±19.37	2.40±8.66	1.20±4.33	NS
(Z)-9- Octadecen-13-olide	2044	5.17±19.30	2.24±5.50	4.73±17.06	NS
Octadecanoic acid ester	2047	1.67±3.97	6.21±11.12	2.10±5.94	NS
Unknown macrolide	2048	1.16±2.28	4.68±6.25	0.53±1.30	NS
Henicosene	2068	53.38±54.05	51.03±29.18	74.85±122.55	NS
Henicosene	2072	11.49±16.42	10.88±8.71	17.30±21.31	NS
Henicosene	2086	5.98±8.64	5.89±5.16	5.41±11.25	NS
Henicosane	2100	1614.47±1165.36	1569.21±536.05	1502.11±1289.54	NS
9-Octadecen-18-olide	2123	0.69±2.08	16.44±29.73	5.79±14.62	NS
18-Octadecanolide	2135	2.62±4.86	4.98±7.66	5.47±13.83	NS
Butyl hexadecenoate	2153	5.63±14.40	1.76±4.62	3.28±8.21	NS
Ethyl oleate	2160	2.22±6.34	5.64±14.76	83.31±245.15	NS
Butyl hexadecanoate	2186	146.79±189.69	73.69±61.43	152.44±235.08	NS
Isopropyl oleate	2188	8.21±26.61	89.25±129.52	82.97±156.19	NS
Docosane	2198	21.18±20.35	26.90±22.85	15.24±16.54	NS
Unknown compound	2250	0±0	1.48±5.32	1.33±2.09	0.039
Tricosene	2270	84.15±83.76	102.54±78.73	180.70±300.84	NS
Tricosene	2275	30.83±54.18	21.16±24.96	16.24±14.90	NS
Isopentyl ester	2280	0.16±0.64	5.89±18.52	6.00±17.52	NS
Isobutyl oleate	2287	19.14±50.51	264.92±578.20	342.60±833.26	NS
Tricosane	2298	140.17±936.38	126.64±107.42	129.47±123.82	NS

Unknown compound	2305	6.85±11.73	2.72±5.56	4.78±12.70	NS
Hexyl hexadecenoate	2353	4.74±17.52	9.25±33.36	2.33±8.39	NS
Butyl oleate	2355	845.44±1032.66	1028.79±939.16	1213.21±1386.81	NS
Butyl octadecadienoate	2355	28.30±65.79	22.08±39.32	28.95±44.74	NS
(*Z*)-3-Hexenyl hexadecanoate	2379	288.69±344.78	297.39±297.88	217.63±310.39	NS
Butyl octadecanoate	2386	58.67±119.39	26.76±30.34	55.17±92.62	NS
Isoprenyl octadec-11-enoate	2436	5.15±18.52	3.17±6.66	12.65±32.87	NS
Unknown compound	2460	2.25±9.26	2.78±10.03	19.24±52.08	NS
(*Z*)-13-Docosen-1-ol	2464	69.50±115.09	55.06±32.28	70.20±114.92	NS
Unknown compound	2466	4.15±17.10	7.10±13.76	0±0	NS
1-Docosanol	2489	408.14±706.49	399.90±295.30	486.75±712.49	NS
Pentacosane	2500	60.51±67.58	53.40±35.11	57.50±55.56	NS
11-Methylpentacosane	2532	20.72±28.09	13.23±15.59	18.77±29.49	NS
Unknown compound	2550	61.86±157.89	42.53±66.66	51.60±104.67	NS
Hexyl octadecenoate and (Z)-3-hexenyl octadecenoate	2557	1398.41±1699.54	1724.97±1123.13	1689.16±2098.32	NS
Hexenyl octadecatrienoate and (Z)-3-hexenyl octadecatrienoate	2561	216.83±278.88	271.38±240.37	323.64±504.26	NS
(Z)-3-Hexenyl octadecanoate	2581	94.71±171.44	88.06±108.84	84.38±148.61	NS
Hexacosane	2600	2.51±6.33	0±0	1.08±3.88	NS

1,3-Docosanediol cyclic dimethylsilyl-derivative	2604	128.18±210.59	208.91±125.60	164.30±355.66	NS
Tetracosenol	2670	601.62±936.38	762.63±414.84	1038.29±1582.15	NS
1-Tetracosanol	2694	226.72±435.64	335.14±296.89	366.72±439.37	NS
Unknown compound	2694	0±0	6.35±22.89	445.39±1236.54	NS
Heptacosane	2700	131.7±210.29	61.09±47.31	55.22±82.73	NS
11-Methylheptacosane	2725	12.99±25.46	6.91±8.16	4.73±9.83	NS
Cholestadiene	2744	0±0	9.56±34.48	18.18±24.10	<0.001
Unknown compound	2746	1.90±7.30	0.12±0.44	1.84±5.05	NS
Tetracosenolide	2749	0±0	13.03±22.58	4.05±14.51	NS
Unknown compound	2750	0±0	5.09±11.07	0±0	NS
Unknown compound	2752	5.41±15.64	0±0	3.51±12.66	NS
Unknown compound	2753	1.92±7.62	1.82±6.56	0±0	NS
Unknown compound	2770	63.97±102.60	332.49±523.57	94.07±153.78	NS
Unknown compound	2783	7.03±20.83	0±0	8.64±31.13	NS
1,3-Tetracosanediol cyclic dimethylsilyl-derivative	2799	225.11±454.13	363.47±270.74	363.61±816.74	NS
Unknown compound	2822	4.08±14.37	0.62±2.22	0±0	NS
Hexacosanal	2829	62.94±125.32	64.13±51.46	158.10±331.50	NS

Table S3: *Pairwise comparisons of dispersion of* H. melpomene *genital compounds when reared on different plants. (Observed p-value below diagonal, permuted p-value above diagonal).*

	P. menispermifolia	*P. platyloba*	*P. vitifolia*
P. menispermifolia		0.002	0.021
P. platyloba	0.007		0.964
P. vitifolia	0.024	0.962	

Table S4: *Androconial compounds identified in* H. melpomene rosina *males fed as adults with (20 individuals) or without pollen (33 individuals) with a mean amount greater than 1 ng in at least one of the treatments. The gas chromatographic retention index (RI) is reported for each compound. Mean amounts (ng) ± standard deviation. No differences were found to be statistically significant.*

Chemical	**RI**	**With pollen**	**Without pollen**
Nonanal	1105	20.02±6.30	28.38±11.46
Benzoic acid	1166	8.82±8.46	7.18±6.99
Naphthalene	1181	2.26±1.36	1.94±2.50
Homovanillyalcohol	1523	1.26±1.71	0.98±2.08
Syringaldehyde	1662	592.20±296.89	501.25±230.87
3,5-Dimethoxy-4-hydroxybenzyl alcohol	1707	3.90±4.40	2.71±3.68
1-(3,5-Dimethoxy-4-hydroxybenzyl)ethanone	1735	4.17±3.54	5.83±5.09
Unknown hydrocarbon	1750	2.32±2.72	2.42±3.42
Ethyl benzoate	1762	1.69±1.19	2.60±2.64
Benzyl benzoate	1766	2.28±2.99	0.67±1.29
1-(4-Hydroxy-3,5-dimethoxyphenyl)-2-propen-1-one	1807	5.23±5.02	6.76±8.41
Methyl 1*H*-indol-3-acetate	1822	4.25±2.99	3.26±3.09
Methyl 1*H*-indol-3-carboxylate	1853	1.47±2.22	1.42±1.79
Unknown compound	1914	1.37±1.61	0.69±0.95
Unknown compound	1930	1.75±2.14	0.79±1.51
Hexadecanoic acid	1960	5.61±5.01	5.44±5.27
(*Z*)-9-Octadecenal	1996	10.69±7.61	9.24±8.44
Octadecanal	2021	688.33±388.77	626.43±327.87
9-Octadecen-1-ol	2040	180.69±7.61	9.24±8.44
Nonadecanal, methyl, branched	2054	5.61±5.01	5.44±5.27
Ethyl 4-hydroxy-3,5- dimethoxybenzoate	2057	16.52±16.30	18.36±19.15
Unknown compound	2057	7.94±4.78	6.90±5.21
Methyloctadecanal	2065	8.13±6.03	7.32±7.39
Henicosadiene	2065	3.31±4.30	2.47±3.94
Methyloctadecanal	2072	6.56±5.21	5.56±5.27
Methyloctadecanal	2077	21.03±14.15	25.57±15.99
Octadecan-1-ol	2082	384.90±286.32	264.82±225.41

(Z)-16-Methyl-9-octadecenol	2092	4.60±6.29	5.91±5.08
Henicosane	2099	10.12±5.72	15.23±7.61
Unknown compound	2103	3.00±7.81	1.95±6.44
Unknown compound	2112	1.62±2.99	0.12±0.71
Unknown alkene or alcohol	2127	27.64±17.93	22.76±14.84
Unknown compound	2132	35.81±36.17	17.46±26.56
Methyloctadecan-1-ol	2138	17.35±11.25	17.10±10.78
Unknown hydrocarbon	2143	1.38±4.23	0.82±3.40
Octadecanoic acid	2160	1.79±3.09	0.54±1.79
Unknown alcohol	2166	1.08±2.24	1.21±2.44
(*Z*)-11-Icosenal	2198	103.12±70.73	105.61±77.15
Icosanal	2224	16.58±9.85	15.82±9.47
(*Z*)-11-Icosenol	2261	768.01±361.82	523.86±343.46
Tricosene	2281	7.76±6.12	6.19±5.45
Fatty acid amide	2325	2.29±30.1	2.80±3.93
Unknown compound	2352	6.03±12.03	3.72±6.66
Unknown amide	2374	2.34±8.40	3.96±19.32
Unknown compound	2392	4.41±6.26	7.88±34.56
(*Z*)-13-Docosenal	2403	44.77±34.60	35.19±30.69
(*Z*)-13-Docosen-1-ol	2464	60.82±52.24	33.23±41.49
Unknown compound	2465	1.78±4.71	4.09±6.46
Pentacosane	2500	10.57±9.64	7.70±6.69

Table S5: *Genital compounds identified in* H. melpomene rosina *males fed as adults with (20 individuals) or without pollen (27 individuals) with a mean amount greater than 1 ng in at least one of the treatments. The gas chromatographic retention index (RI) is reported for each compound. Mean amounts (ng) ± standard deviation. No differences were found to be statistically significant.*

Chemical	RI	With pollen	Without pollen
β-Myrcene	990	9.39±7.27	14.87±11.98
(*Z*)- β-Ocimene	1030	881.98±3725.80	971.89±4438.55
(*E*)- β-Ocimene	1054	21179.02±10155.09	30487.62±18517.35
Nonanal	1105	3.61±2.43	4.34±5.42
Alloocimene	1129	21.45±15.80	39.62±39.16
(4*E*,6*Z*)-2,6-Dimethyl-2,4,6-octatriene	1140	10.10±11.40	13.04±18.50
2-sec-Butyl-3-methoxypyrazine	1173	25.95±20.86	25.81±20.67
2-Methoxy-3-isobutylpyrazine	1181	10.61±8.27	9.59±5.70
Napthalene	1181	1.32±1.56	4.34±7.71
Methyl salicylate	1189	4.51±9.46	5.35±10.85
Dihydroedulan II	1290	8.78±11.13	15.20±14.38
Unknown sesquiterpene	1384	6.30±11.42	12.43±23.95
Unknown compound	1396	16.07±13.93	26.85±20.84
β-Caryophyllene	1417	3.88±10.48	8.91±29.30
Unknown compound	1433	0.84±2.05	2.36±4.05
11-Dodecanolide	1465	6.73±9.05	8.40±9.69
14-Tetradecanolide	1733	11.74±24.63	4.47±11.99
Ethyl benzoate	1762	1.44±1.62	1.57±2.09
Hexadecenolide	1845	5.53±17.19	9.55±24.97
Unknown compound	1859	0.79±3.19	1.21±5.21
Nonadecane	1899	1.61±3.35	1.42±3.36
Hexadecanoic acid	1960	15.85±19.88	17.87±29.71
(*Z*)-9-Octadecenal	1996	7.50±10.06	6.10±12.93
Icosane	1998	18.11±19.35	19.47±19.80
Octadecadienolide	2035	3.31±4.86	1.31±2.96
(Z)-9- Octadecen -13-olide	2044	7.16±17.20	29.61±71.06
Octadecanoic acid ester	2047	8.75±17.77	14.61±35.16
Unknown macrolide	2048	3.77±8.21	4.27±11.31

Henicosene	2068	84.17±89.07	82.57±111.40
Henicosene	2072	40.57±27.09	47.44±29.12
Henicosene	2086	14.82±16.54	12.33±14.01
Henicosane	2100	1673.83±1185.53	2040.62±1566.25
9-Octadecen-18-olide	2123	11.57±23.92	17.70±42.81
18-Octadecanolide	2135	13.74±18.73	25.94±35.28
Butyl hexadecenoate	2153	25.92±57.62	15.67±49.62
Ethyl oleate	2160	147.50±457.41	10.19±35.77
Butyl hexadecanoate	2186	290.97±395.89	263.52±358.96
Isopropyl oleate	2188	66.81±130.57	202.02±428.01
Docosane	2198	26.73±24.06	26.97±27.21
Icosanal	2228	4.45±4.07	10.26±12.09
Isoprenyl palmitate	2254	3.10±6.48	4.32±17.03
Fatty acid amide	2325	1.07±2.70	3.92±9.83
Unknown compound	2250	1.97±2.93	2.73±4.94
Tricosene	2270	127.52±118.14	174.15±312.73
Tricosene	2275	53.20±33.49	239.55±784.00
Isopentyl ester	2280	2.93±6.25	34.07±115.31
Isobutyl oleate	2287	272.35±494.86	456.07±1058.14
Tricosane	2298	151.85±140.12	145.52±150.42
Unknown compound	2305	25.00±50.08	17.72±30.82
Butyl oleate	2355	2214.19±2368.51	2678.85±2991.26
Butyl octadecadienoate	2355	63.55±84.53	117.52±197.94
(*Z*)-3-Hexenyl hexadecanoate	2379	346.46±490.37	280.53±356.56
Butyl octadecanoate	2386	94.73±130.80	103.26±153.70
Tetracosane	2400	3.31±4.86	1.31±2.96
19-Methyleicosyl acetate	2422	5.31±11.84	6.75±14.63
Unknown aldehyde	2422	9.51±7.39	7.60±9.96
Isoprenyl octadec-11-enoate	2436	30.34±51.83	66.75±177.60
(*Z*)-13-Docosen-1-ol	2464	89.49±117.01	173.83±219.21
1-Docosanol	2489	496.28±431.58	881.27±900.58
Pentacosane	2500	137.90±141.25	100.48±84.41
11-Methylpentacosane	2532	66.03±83.60	32.51±38.54
Hexyl octadecadienoate	2544	33.98±130.16	8.12±40.30
Hexyl octadecenoate and (*Z*)-3-hexenyl octadecenoate	2557	2058.15±2294.58	3106.82±3582.23

Hexenyl octadecatrienoate and (*Z*)-3-hexenyl octadecatrienoate	2561	309.53±337.68	364.39±579.66
(*Z*)-3-Hexenyl octadecanoate	2581	140.28±188.85	161.72±240.14
Hexyl octadecanoate	2590	77.96±152.82	198.52±347.46
Hexacosane	2600	29.70±30.06	13.90±28.32
1,3-Docosanediol cyclic dimethylsilyl-derivative	2604	119.97±115.65	327.36±499.90
Tetracosenol	2670	786.60±733.25	1626.51±1667.51
1-Tetracosanol	2694	491.61±567.03	1108.74±1589.39
Heptacosane	2700	149.12±88.41	203.20±138.24
11-Methylheptacosane	2725	29.63±30.91	51.28±39.67
Unknown compound	2746	7.77±16.27	0±0
Unknown compound	2770	201.71±348.54	174.97±163.95
Unknown compound	2783	12.57±22.04	40.47±57.57
1,3-Tetracosanediol cyclic dimethylsilyl-derivative	2799	153.20±166.52	405.72±479.21
Octacosane	2800	11.58±21.23	7.52±15.09
Hexacosanal	2829	65.41±66.25	247.52±388.10

Chapter 4

Species specificity and intraspecific variation in the chemical profiles of *Heliconius* butterflies across a large geographic range

Abstract

In many animals, mate choice is important for the maintenance of reproductive isolation between species. Traits important for mate choice and behavioural isolation are predicted to be under strong stabilising selection within species, however such traits can also exhibit variation at the population level driven by neutral and adaptive evolutionary processes. Here, we describe patterns of divergence among androconial and genital chemical profiles at inter- and intra-specific levels in mimetic *Heliconius* butterflies. Most variation in chemical bouquets was found between species, but there were also quantitative differences at the population level. We found a strong correlation between interspecific chemical and genetic divergence, but this correlation varied in intraspecific comparisons. We identified 'indicator' compounds characteristic of particular species that included compounds already known to elicit a behavioural and electrophysiological response, suggesting an approach for identification of candidate compounds for future behavioural studies in novel systems. Overall, the strong signal of species identity suggests a role for these compounds in species recognition, but with additional potentially neutral variation at the population level.

Introduction

Reproductive isolation between lineages is important for the maintenance of species diversity (Coyne and Orr, 2004). In many animals, mate choice provides a strong pre-mating barrier, maintaining reproductive isolation (Friberg et al., 2008; Gray and Cade, 2000; Martin and Mendelson, 2016; Nagel and Schluter, 1998; Ready et al., 2006; Seehausen et al., 2008; Selz et al., 2014). Closely related species often differ in traits important for mate choice, with individuals displaying a preference for conspecific phenotypes (Jiggins et al., 2001; Mas and Jallon, 2005; Ryan and Guerra, 2014; Saveer et al., 2014; Yildizhan et al., 2009). These traits are predicted to show strong species-specific differences (Gerhardt, 1982), and typically should be subject to stabilising selection which can act to decrease intraspecific phenotypic variation (Butlin et al., 1985; Pfennig, 1998; Ptacek, 2000). As a consequence, we would expect to find little trait variability, or at least certain features to be invariant, across species geographic ranges (Benedict and Bowie, 2009; Ferreira and Ferguson, 2002; McPeek et al., 2011; Weber et al., 2016). However, these traits can also exhibit variation both within and between populations of the same species, either due to genetic drift and/or varying selective regimes across their ranges (Bolnick and Kirkpatrick, 2012; Ryan et al., 1996; Ryan and Guerra, 2014; Ryan and Rand, 1993).

Signals important for behavioural isolation could arise from the divergence of traits used in intraspecific communication between populations (Johansson and Jones, 2007; Mendelson and Shaw, 2012; Ryan and Rand, 1993; Smadja and Butlin, 2008). Signal divergence can be driven by various factors, both neutral and adaptive, usually involving multiple evolutionary forces (Leonhardt et al., 2013; Sun et al., 2013). A positive correlation between genetic distance and phenotypic variation is consistent with stochastic processes, such as genetic drift, playing a prominent role (Irwin et al., 2008). In contrast, a lack of correlation between phenotypic and genetic divergence may suggest that selection is shaping the phenotypic variation, perhaps driving divergence in different directions in each population (Campbell et al., 2010; Conrad et al., 2018; Hankison and Ptacek, 2008; Mullen et al., 2009).

Chemical compounds, such as sex pheromones, mediate intraspecific communication in many systems (Wyatt, 2014, 2003). The role of chemical signalling in behavioural isolation is also well established, especially among moth species (Löfstedt, 1993; Smadja and Butlin, 2008). Pheromone evolution requires changes in both the detection of pheromone by the receiver, and the production of pheromone by the sender. Due to this coordination between detection and production, these pheromone blends are traditionally regarded as being under stabilising selection towards a species stereotype (Löfstedt 1993). Nonetheless, even when species-specific characteristics are present, chemical composition can exhibit intraspecific variation, with both qualitative and quantitative differences found across a species range (Carde and Allison, 2016).

Studies of *Heliconius* butterflies have contributed to our understanding of adaptation and speciation (Jiggins, 2017, 2008; Merrill et al., 2015). Despite the reliance of this group on visual cues for mating (Bybee et al., 2012; Finkbeiner et al., 2017; Jiggins et al., 2001; Merrill et al., 2014; Sánchez et al., 2015), it has long been suggested that male pheromones also play a role in pre-mating barriers (Jiggins, 2008; Merrill et al., 2015), but so far only a few species have been studied. Behavioural experiments reveal that chemical signalling in *Heliconius erato*, *H. melpomene* and *H. timareta* is important for female mate choice (Darragh et al., 2017; Mérot et al., 2015). Electroantennographic studies have shown that *Heliconius cydno* and *H. melpomene* respond to both con- and heterospecific androconial chemical bouquets (Byers et al., 2019), and have identified an individual compound that is electrophysiologically and behaviourally active. Furthermore, studies of *H. cydno*, *H. doris*, *H. hecale*, *H. ismenius*, *H. melpomene*, *H. pardalinus*, *H. sara* and *H. timareta* from Panama, Colombia, Ecuador, and Peru, found that major compounds differ between species (Mann et al., 2017; Mérot et al., 2015), suggesting a potential role in reproductive isolation.

The role of chemical signalling is likely to be especially important in co-mimics, where visual signals alone are not sufficient to identify conspecifics (Estrada and Jiggins, 2008; Giraldo et al., 2008; Mérot et al., 2013; Sánchez et al., 2015). In contrast, chemical compounds could be part of a multimodal aposematic warning signal (Rojas et al., 2018; Rothschild, 1961), with some tentative evidence that co-mimics exhibit similar chemical bouquets to aid recognition by predators (Mann et al., 2017).

Here, we describe the chemical profiles of seven species of *Heliconius* from over 250 individuals collected across the Neotropics. We focus on the co-mimetic species *H. melpomene* and *H. erato* that are distributed widely across the Neotropics and analysed both wing androconial and genital compounds of male butterflies. We hypothesise that compounds found consistently across the geographic range of a species are likely to be behaviourally active compounds, important for mate choice. We use *H. melpomene* as a test-species due to the availability of behavioural and electrophysiological data to investigate this approach, by evaluating consistency in compound blends across different localities.

The extensive dataset analysed here allows us to test evolutionary hypotheses, as well as identifying interesting candidate compounds for future behavioural studies. As well as interspecific variation, we also investigated intraspecific variation in chemical profiles of *H. melpomene* and *H. erato*. In both inter- and intraspecific datasets, we correlated chemical profile data with both geographic and genetic distances. Furthermore, to investigate if the chemical compounds are part of the aposematic co-mimicry signal we sampled two different mimicry rings in western Ecuador and Panama.

Methods

Sampling

Between February 2016 and August 2017 wild males of *Heliconius cydno*, *H. elevatus*, *H. eleuchia*, *H. erato*, *H. melpomene*, *H. sapho* and *H. timareta* were collected with hand nets from twelve localities. Between two and fifteen males were chemically analysed per population (Fig. 1, Table S1), and one representative from each subspecies of *H. erato* and *H. melpomene* were used for whole-genome sequencing (Table S2). We follow the latest *Heliconius* taxonomy (Lamas and Jiggins, 2017).

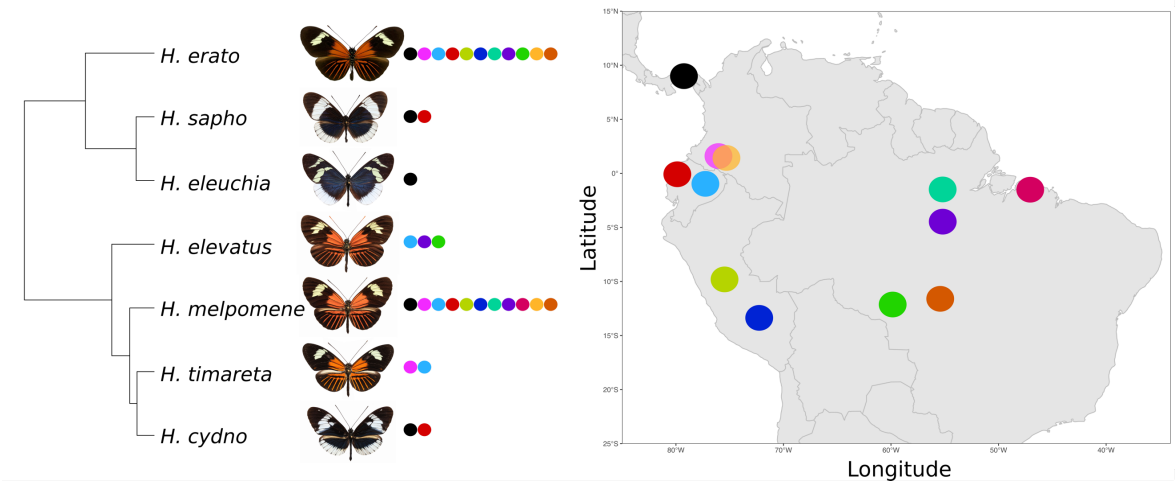


Figure 1. *Map indicating species collected from twelve localities across the Neotropics. See Table S1 for sample numbers. The phylogeny was previously published by Kozak et al., (2015).*

Extraction and chemical analysis of tissues

The androconial region of the wing, previously described as the grey-brown overlapping region of the hindwing (Darragh et al., 2017), as well as the genitalia, were dissected for analysis immediately after collection. For chemical extraction, the tissue was soaked in 200µl dichloromethane containing 200ng 2-tetradecyl acetate (internal standard) in 2ml glass vials with PTFE-coated caps (Agilent, Santa Clara, USA) for one hour. The solvent was then transferred to new vials, maintained cool in the field and stored at -20°C upon return. Androconial samples were evaporated to a reduced volume at room temperature prior to analysis. Extracts were analysed by GC/MS using an Agilent model 5977 mass-selective detector connected to an Agilent GC model 7890B and equipped with an Agilent ALS 7693 autosampler. HP-5MS fused silica capillary columns (Agilent, Santa Clara, USA, 30 m × 0.25 mm, 0.25 µm) were used. Injection was performed in splitless mode (250°C injector temperature) with helium as the carrier gas (constant flow of 1.2 ml/min). The temperature programme started at 50°C, was held for 5 minutes, and then rose at a rate of 5°C/minute to 320°C, before being held at 320°C for 5 minutes. Components were identified by comparison of mass spectra and gas chromatographic retention index with those of authentic reference samples and also by analysis of mass spectra. Components were quantified using 2-tetradecyl acetate as an internal standard. Only compounds eluting earlier than hexacosane were analysed in androconial samples, and those earlier than nonacosane in genital samples (Darragh et al. 2017). We globally removed compounds that were not found in at least half of all individuals from a given population.

DNA extraction and library preparation

We used a representative individual from each subspecies of *H. erato* and *H. melpomene* from across their range. Individuals were genotyped with medium to high coverage whole genome sequencing. We used two sequencing approaches. Genomic DNA of individuals whose ID start with SR or KK (C. Kozak collection, n=14) was extracted from thorax tissue with Qiagen MagAttract beads. The KAPA Biosystems kit was used to prepare paired-end 2x150 base pair libraries with inserts of 50-200bp after SPRI size-selection. Libraries were quality-controlled using the Agilent 2100 Bioanalyzer and

indexed with the KAPA Single-Indexed Adapter Kit. Libraries were sequenced on the Illumina HiSeq4000 platform by Novogene, Tianjin, People's Republic of China. For the remainder newly sequenced individuals (n=6, starting with CAM or 14N, Cambridge and N.Nadeau collections, respectively), we extracted genomic DNA with Qiagen DNeasy kits (Qiagen) from thorax tissue. The libraries for these individuals were TruSeq Nano, gel-free libraries and were sequenced on Illumina HiSeq 2500 platform (v4 chemistry) by Novogene (Hong Kong). Whole-genomes for four individuals were obtained from public databases. Individual information can be found in Table S2.

Calculation of genetic and geographic distance matrices

To explore genetic distance among the studied *H. erato* (n=12) and *H. melpomene* (n=13) populations, we computed whole-genome genetic covariance matrices and performed MDS for each species separately. A whole-genome sequence from a representative individual from each population was used (Table S2). Genotypes were inferred from reads mapped to the *H. melpomene* (v2.5) and *H. erato demophoon* genome scaffolds (Challis et al., 2016; Davey et al., 2017; Heliconius Genome Consortium, 2012; Van Belleghem et al., 2017) with bwa v0.7.15 (Li and Durbin, 2009). We computed a whole-genome pairwise identical by state (IBS) matrix with a random sampled read from each position in the genome, implemented in ANGSD v0.912 (Korneliussen et al., 2014) (angsd -bam bam.path.list -minMapQ 30 -minQ 20 -GL 2 -doMajorMinor 1 -doMaf 1 -SNP_pval 2e-6 -doIBS 1 -doCounts 1 -doCov 1 -makeMatrix 1 -minMaf 0.05).

An interspecific genetic distance matrix was constructed using the function "cophenetic.phylo" from the ape package (Paradis and Schliep, 2018) with a previously published phylogeny (Kozak et al., 2015). Geographic distance matrices were created by inputting the co-ordinates of collection localities into the function "distm" in the geosphere package to calculate the Haversine great-circle-distance between points (Hijmans, 2017).

Statistical analyses: Inter- and intra-specific indicator compounds

We carried out indicator analysis using the indicspecies package (Cáceres and Legendre, 2009). Groupings are decided a priori (in this case species or population) and compounds are determined which act as indicators of these groups. The best indicators are those which are only found in a single group (specificity) and all group members possess the compound (coverage); such a compound would have an indicator value of 1. The specificity of a compound is calculated based on the amount of compound found in each individual, whilst the coverage considers only presence or absence of the compound. We used the function "indicators" to investigate both which single compounds and combinations of compounds best predict group membership. We used the function "pruneindicators" to find the single or combinations of compounds which had the highest indicator values.

Statistical analyses: Variation in chemical profiles

Divergence in chemical profiles across species and populations was estimated with non-metric multidimensional scaling (NMDS) ordination in three dimensions, based on a Bray-Curtis similarity matrix. We used the "metaMDS" function in the *vegan* package version 2.5-1 (Oksanen et al. 2017), and visualised the NMDS using the *ade4* package (Dray and Dufour 2007).

We assessed the relative importance of relevant factors in driving the variation in chemical profiles with multivariate statistical analyses. These factors included: species identity, geographic region and individual locality. We excluded subspecies as a factor because, in *Heliconius*, these are determined based on their, sometimes very subtle, difference in wing colour pattern, with extensive gene flow across the genome between subspecies (Van Belleghem et al., 2017). It is therefore more biologically relevant to include locality in the model, to account for genetic drift between subspecies, and since locality and subspecies are highly correlated we cannot include both. To compare overall variation in chemical composition between groups, we carried out PERMANOVA (permutational multivariate analysis of variance) testing based on a Bray-Curtis distance matrix, using the "adonis2" function in the vegan package with 1000 permutations. We

investigated each term in the model sequentially, starting with species identity, the main clustering factor found from visualisation with NMDS, followed by geographic region (Panama vs Western Andes vs Eastern Andes vs Amazon), and finally individual collecting localities. Model goodness of fit was evaluated by Akaike information criterion (AIC). In general, we chose the model with the lowest AIC value, however, if two models were within two AIC of each other we chose the simplest model as the best fit (Table S3). We followed these PERMANOVA tests with post hoc pair-wise testing using the function "pairwise.perm.manova" in the RVAidememoie package, with Bonferroni correction, to identify which grouping factors were significantly different (Hervé, 2018). We repeated the PERMANOVA analysis within species, in *H. erato* and *H. melpomene*, to investigate fine-scale intraspecific geographic patterns. In the within species analysis we included geographic region (Panama vs Western Andes vs Eastern Andes vs Amazon), and individual collecting localities as the two factors.

One issue with distance-based analyses such as PERMANOVA is that differences in dispersion between groups can be confounded with differences in location (Warton et al., 2012). To confirm these analyses and account for this issue, we implemented multivariate generalised linear models using the function "ManyGLM" from the mvabund package (Wang et al., 2012). We modelled the data using a negative binomial distribution, which we found to be appropriate through examination of residual plots. For interspecific analyses we included species, region, and locality nested within region in the model. For intraspecific analyses we included region and locality nested within region. The "ManyGLM" function fits models to each chemical compound, summing the test statistics to give a multivariate test statistic known as Sum-of-LR. This statistic can be tested for significance using resampling methods. We carried out backwards elimination and compared the fit of models by using the "anova.manyglm" function with a likelihood ratio test (Table S4). We can also determine which compounds are driving between-group differences by looking at the individual contribution of each compound to the Sum-of-LR, with p-values adjusted for multiple testing using the "adjust" option.

Phylogenetic and geographic distance

Shared ancestry can explain part of the variation in a species' chemical profile. Using the interspecific genetic distance matrix calculated above, we tested for a correlation between phylogenetic distance and chemical profile divergence. We carried out partial Mantel tests, controlling for geographic distance, using the vegan package (Oksanen et al., 2017). To investigate the role of geographic distance in chemical profile divergence, we compared geographic and chemical distances matrices, controlling for genetic distance, with partial Mantel tests. To visualise the species phylogeny (Kozak et al., 2015) we used the "plot.phylo" function from the ape package (Paradis and Schliep, 2018).

Genomic and chemical distance within species

We calculated intraspecific genetic distances using genome sequences from 11 *H. erato* and 13 *H. melpomene* populations. We visualised genetic distances in two dimensions using MDS with the function "cmdscale". We tested for a correlation between intraspecific genetic distance and chemical profile divergence with partial Mantel tests, controlling for geographic distance, using the vegan package (Oksanen et al., 2017). Hybrids between populations of the same species were excluded from this analysis (Table S2). We also used partial Mantel tests to investigate the role of geographic distance, while controlling for genetic distance.

Co-mimics and similarity of chemical profiles

We used samples of two mimicry rings from two localities, Panama and western Ecuador. *H. melpomene* and *H. erato* form one mimicry ring, whilst *H. cydno* and *H. sapho* form another, with the addition of *H. eleuchia* in western Ecuador (Fig. 1). We visualised these samples but did not carry out statistical analyses due to the psuedoreplication caused by the similar of individuals within a species. More species comparisons would be needed for further analysis.

All statistical analyses were performed with R version 3.5.1 (R Core Team, 2018). Figures were made using a palette of colours optimized for colour-blindness (Wong, 2011). We used ggplot2 for violin and boxplots (Wickham, 2009). Sequencing data is available from ENA under accession number PRJEB35570. GCMS chromatograms, other data, and R scripts used for analysis are available from Open Science Framework (OSF) (https://osf.io/28yfk/).

Results

Chemical compounds in androconia and genitals

We sampled 252 androconia and 275 genitals across 42 populations of seven species, and identified 349 compounds in the genitals and 157 in the androconia (Table S5, S6). Of the total number of androconial compounds, 38% are fatty acid derivatives, 20% aromatics, 10% terpenoids, 1% macrolides, <1% lactones and 31% unknown or unidentified compounds. Of the genital compounds, 17% are fatty acid derivatives, 7% aromatics, 10% terpenoids, 1% lactones, 12% macrolides and 44% unknown or unidentified compounds. The main difference is that there are more macrolides in the genitals than in androconia.

Heliconius species varied considerably in the amount and number of compounds (Fig. 2). Between species there was variation in the number of compounds per individual, and the overall amount of compounds detected (Table S5, S6). For the androconia, *H. eleuchia* had the fewest compounds (13±5), and *H. melpomene* the highest (32±11) (mean ± standard deviation). *H. sapho* had the lowest total amount of androconial compounds at 1,300±803 ng, and *H. melpomene* the highest at 7,254±8,242 ng. The species with the fewest genital compounds was *H. sapho* with 32±7, and the highest *H. cydno* with 102±21. *H. sapho* also had the lowest total amount of genital compounds at 6,642±3,975 ng, and *H. cydno* the highest at 91,167±67,122 ng. These values are within the same order of magnitude as expected from previous work on male sex pheromones in the butterfly *Bicyclus anyana* (Nieberding et al., 2012; van Bergen et al., 2013). Using *H. erato* as an example, the androconial bouquet is 0.00002%, and genital bouquet 0.0007% of total body weight (Montgomery et al., 2016). In general, a higher number of

compounds and total amount of compounds is found in the genitals than in the androconial patches of *Heliconius* wings.

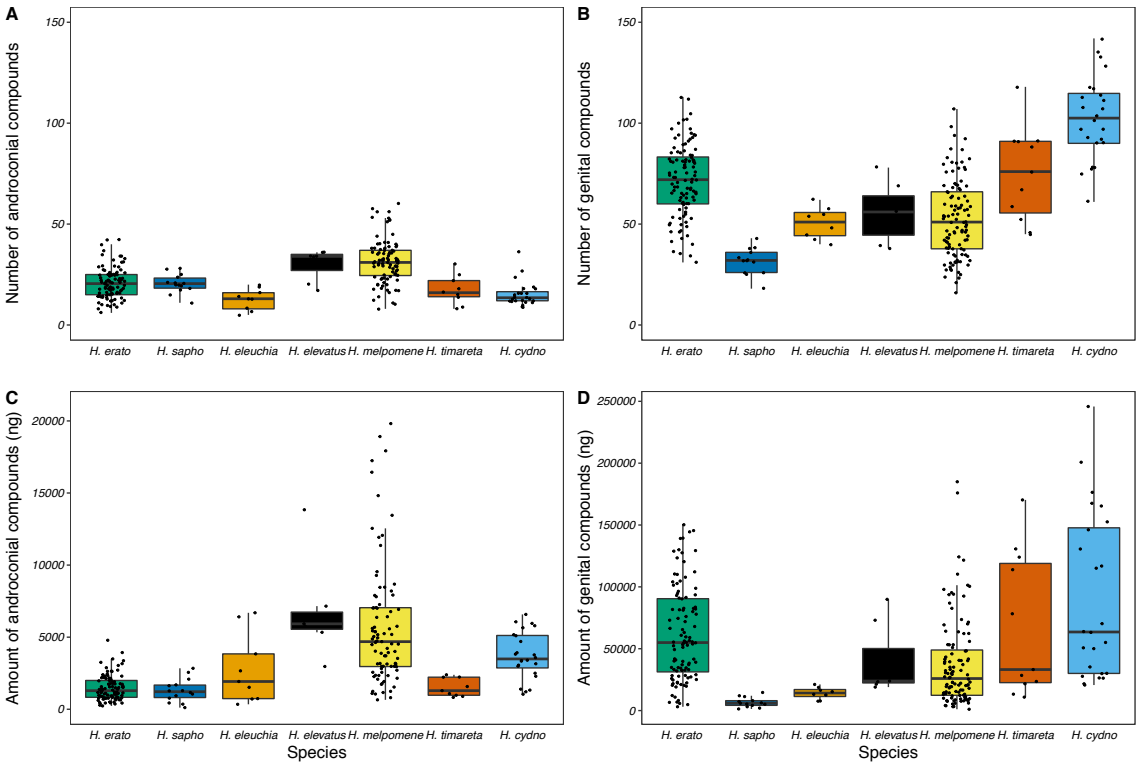


Figure 2. *Composition of androconial and genital bouquets across seven* Heliconius *species. Species significantly differ in: (A) number of androconial compounds (ANOVA, $F_{6,245}$=21.54, p<0.001), (B) number of genital compounds (ANOVA, $F_{6,268}$=36.15, p<0.001), (C) amount of androconial compounds (ANOVA, $F_{6,245}$=11.55, p<0.001), (D) amount of genital compounds (ANOVA, $F_{6,268}$=11.62, p<0.001). Four outlier individuals were removed from C.*

Are there species-specific chemical compounds?

In order to identify candidate species recognition pheromones, we examined our data for species specific compounds using indicator analysis. In most species that we examined there were single androconial compounds that were strong indicators of species identity (Table 1). For example, geranylgeranylacetone was found only in *H. erato* and was consistently present across samples of this species. Similarily, octadecanal, a compound previously shown to be electrophysiologically active (Byers et al. 2019), was found almost exclusively in *H. melpomene* (specificity=0.999). *H. cydno* and *H. eleuchia* had the weakest indicator scores; in *H. cydno* because the best indicator compound was not found in all individuals examined (coverage= 0.667), and in *H. eleuchia* because the best indicator compound was also found in other species (specificity= 0.747). There were similarly species-specific genital compounds in all species except *H. sapho* and *H. timareta*, where a combination of two compounds was the best predictor (Table 2). Similar to the androconia, in *H. melpomene*, the best indicator compound for genitalia has known biological activity, in this case the anti-aphrodisiac, (*E*)-β-ocimene (Schulz et al. 2008). For *H. erato* we identified a terpene ester which is only found in *H. erato* individuals and no other species. Other terpene esters were also almost perfect indicator compounds of *H. erato*.

Table 1: *Androconial compounds which are the best indicators of species identity. A is a measure of species specificity of the compounds, B is a measure of species coverage, and sqrtIV is the indicator value which considers both A and B and ranges from 0 (compound not present in any individuals of that species) to 1 (compound only present in that species, and present in all individuals).*

Species/compound	**A: specificity**	**B: coverage**	**sqrtIV**
H. cydno			
Unknown aromatic (RI=2130)	1	0.667	0.816
H. eleuchia			
Hexahydrofarnesylacetone	0.747	1	0.864
H. elevatus			
Homovanillylalcohol	0.912	1	0.955
H. erato			
Geranylgeranylacetone	1	1	1
H. melpomene			
Octadecanal	0.999	1	1
H. sapho			
Methyl 4-hydroxy-3-methoxybenzoate	0.866	1	0.931
H. timareta			
5-Decanolide	1	0.889	0.943

Table 2: *Genital compounds which are the best indicators of species identity. A, B, and sqrtIV as in Table 1.*

Species/compound	A: specificity	B: coverage	sqrtIV
H. cydno			
Unknown ester (RI=1390)	0.999	1	0.999
H. eleuchia			
Unknown macrolide RI=1878	0.969	1	0.984
H. elevatus			
Icosenol	0.908	1	0.953
H. erato			
Unknown terpene ester (RI=2494)	1	1	1
H. melpomene			
(*E*)-β-Ocimene	0.865	1	0.930
H. sapho			
(*Z*)-3-Hexenyl isobutyrate & unknown (RI=1691)	0.957	0.923	0.940
H. timareta			
Butyl oleate & (*Z*)-9-octadecen-13-olide	0.915	1	0.956

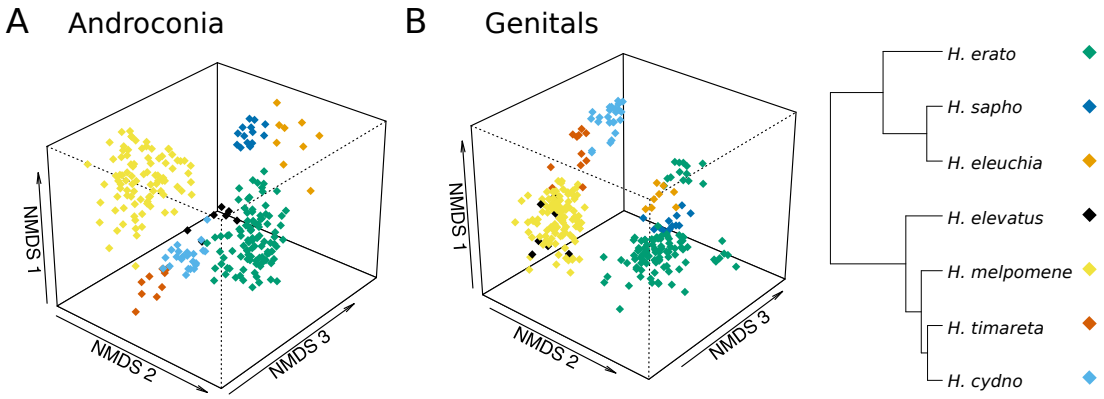


Figure 3. *NMDS (non-metric multidimensional scaling) plot illustrating in three dimensions the variation in chemical compounds of male* Heliconius *of different species. A) Androconial compound bouquets differ significantly between species. Stress=0.155. B) Genital bouquets also differ significantly between species. Stress=0.121.*

What factors affect interspecific variation in chemical profiles?

Our sampling allowed us to investigate how variation in chemical composition is partitioned within and between species, and determine the extent to which chemistry is a species-diagnostic trait. Visualisation of the chemical profiles reveals that individuals mostly group by species for both androconial and genital chemical bouquets (Fig. 3). Species significantly differ in their androconial bouquet, with species identity accounting for 58% of the overall variation in chemical profiles (PERMANOVA, Species, $F_{6,251}$=72.16, p<0.001). All pairwise comparisons of species are significant (Table S7). A further 4% of variation can be explained by region (Amazon/Eastern Andes/Western Andes/Panama), and 3% by locality nested within region (PERMANOVA, Region, $F_{3,251}$=9.96, p<0.001; (Region/Locality), $F_{8,251}$=2.65, p<0.001). Finally, 4% of variation is explained by an interaction between species and region (PERMANOVA, Species*Region $F_{6,251}$=4.82, p<0.001;).

The results were similar for the genital bouquets, with species identity accounting for 51% of the variation in chemical profiles (PERMANOVA, Species, $F_{6,274}$=59.81, p<0.001). All pairwise comparisons are significant apart from *H. elevatus* and *H. melpomene* (Table S8). A further 5% of variation can be explained by region (Amazon/Eastern Andes/Western Andes/Panama), and 3% by locality nested within region (PERMANOVA, Region, $F_{3,274}$=12.43, p<0.001; (Region/Locality), $F_{8,274}$=2.92, p<0.001). Finally, 6% of variation is explained by an interaction between species and region (PERMANOVA, Species*Region $F_{6,274}$=6.52, p<0.001). For both androconial and genital chemical profiles, most variation is explained by species identity, rather than geographic location, as confirmed by ManyGLM (Tables S9, S10). We also confirmed this by comparison of within and between species and locality Bray-Curtis distances (Figure S1, S2, supplementary results).

Does phylogenetic distance explain chemical profile divergence?

Using whole-genome sequence data, we explored the degree to which variation between species can be explained by geographic and genetic distance among the samples. We carried out partial Mantel tests to investigate the correlation between two variables whilst controlling for a third variable. When controlling for geographic distance, genetic divergence is strongly correlated with both androconial and genital chemical divergence (partial Mantel test, androconia, r=0.7871, p=0.001; genitals, r=0.6936, p=0.001). When controlling for genetic distance, geographic distance is significantly but weakly correlated with androconial and genital chemical divergence (partial Mantel test, androconia, r=0.072, p=0.001; genitals, r=0.046, p=0.007).

Do we find population-specific chemical compounds?

We used an indicator analysis to search for compounds unique to specific populations of *H. erato* and *H. melpomene*. Most intraspecific differences are due to quantitative rather than qualitative differences between populations, perhaps explaining why many population indicators were weak as they are also found in other regions at different amounts (Table S11, S12). The only exception is *H. e. cyrbia* (Western Ecuador) that has many genital compounds unique to this region (Table S11).

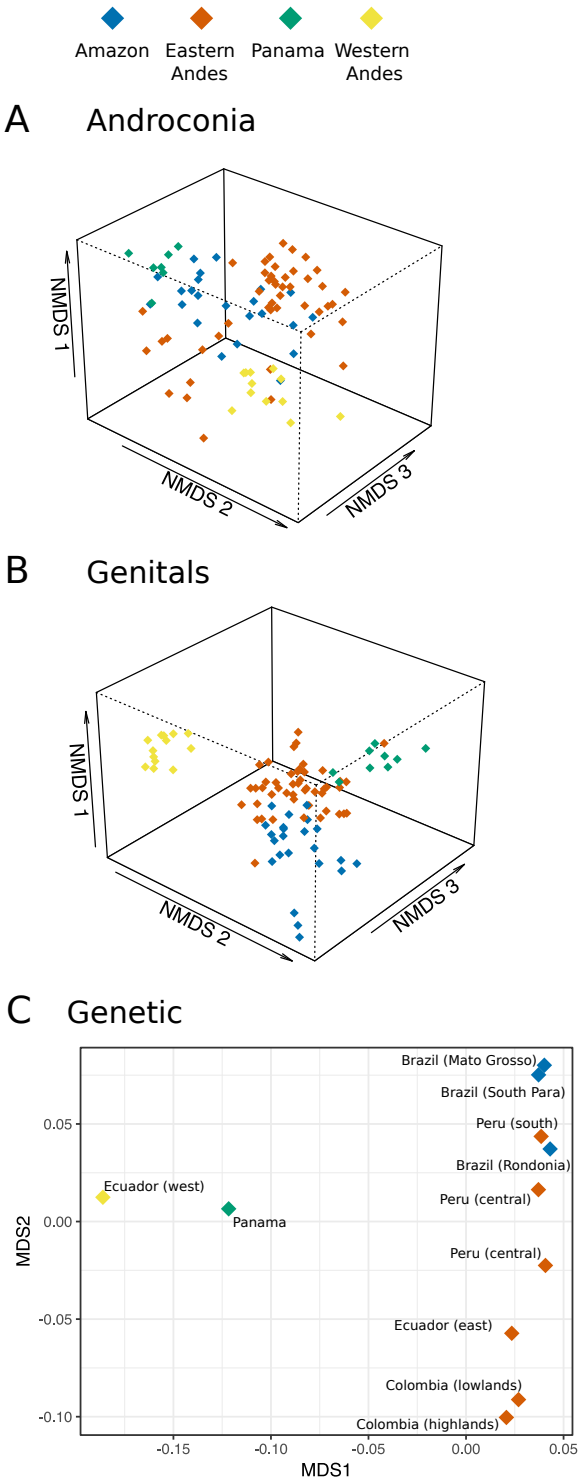


Figure 4. *Plots of androconial, genital and genetic distance between* H. erato *populations. A) NMDS (non-metric multidimensional scaling) plot illustrating in three dimensions the variation in androconial chemical compounds. Stress=0.174. B) NMDS plot illustrating in three dimensions the variation in genital chemical compounds. Stress=0.118. C) MDS plot illustrating in two dimensions the genetic distance between populations of* H. erato*.*

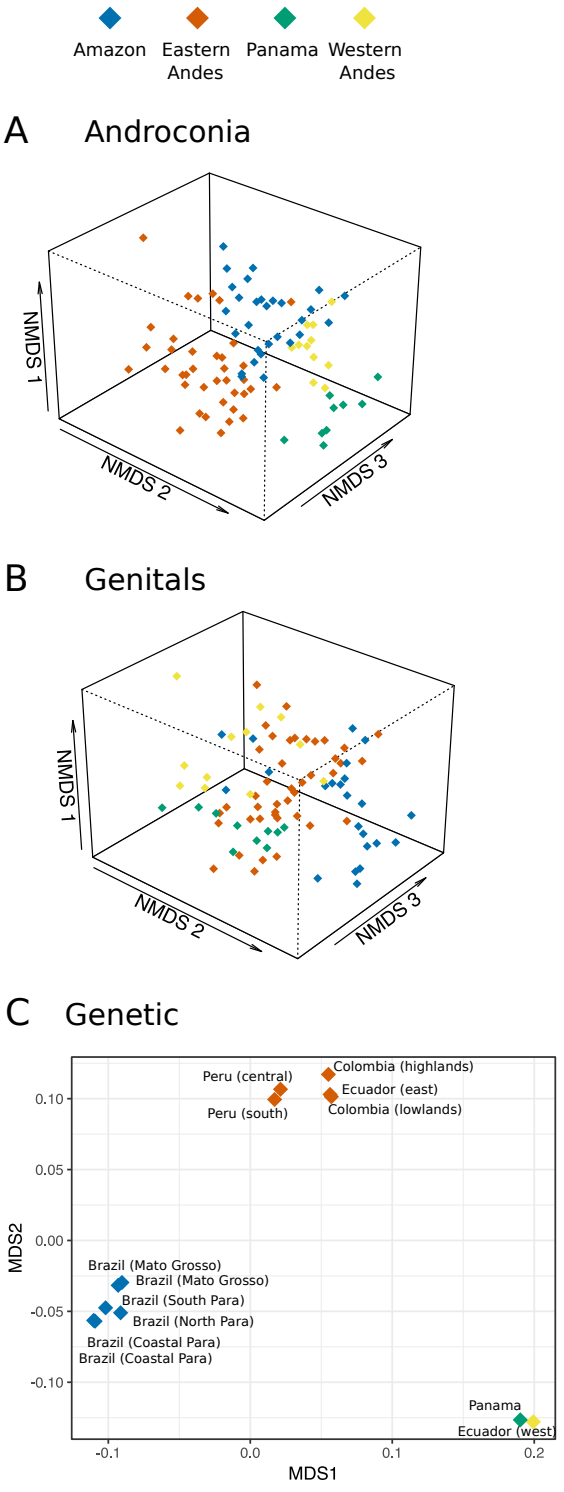


Figure 5. *Plots of androconial, genital and genetic distance between* H. melpomene *populations. A) NMDS (non-metric multidimensional scaling) plot illustrating in three dimensions the variation in androconial chemical compounds. Stress=0.151. B) NMDS plot illustrating in three dimensions the variation in genital chemical compounds. Stress=0.161. C) MDS plot illustrating in two dimensions the genetic distance between populations of* H. melpomene.

What factors affect intraspecific variation in chemical profiles of H. erato *and* H. melpomene*?*

We also wanted to determine the sources of variation within species using our broad sampling of populations across the ranges of *H. erato* and *H. melpomene*. For *H. erato* there was a strong grouping of individuals by region (Fig. 4), with 27% of variation in androconial profiles being explained by region and 11% by locality nested within region (PERMANOVA, Region $F_{3,87}$=11.16, p<0.001, Locality $F_{6,87}$=2.35, p<0.001). All four regions are significantly different from each other (Pairwise permutation MANOVAs, p<0.01). For *H. erato* genital compounds, 37% of variation is explained by region, and 11% by locality nested within region (PERMANOVA, Region $F_{3,91}$=19.01, p<0.001, Locality $F_{6,91}$=2.83, p<0.01). All four regions are significantly different from each other (Pairwise permutation MANOVAs, p<0.05).

These geographic differences in chemical profiles are not as strong in *H. melpomene* (Fig. 5). For *H. melpomene* androconial compounds, the best model only includes region, not locality, with 18% of variation is explained by region (PERMANOVA, Region $F_{3,86}$=6.73, p<0.01). The West Andes subspecies (*H. m. cythera*) is not significantly different from either East Andes (multiple populations) or Panama (*H. m. rosina*) (Pairwise permutation MANOVAs, p=0.072), however, the other comparisons are significantly different (Pairwise permutation MANOVAs, p<0.05). For *H. melpomene* genital compounds, 20% of variation is explained by region, and 12% by locality nested within region (PERMANOVA, Region $F_{3,103}$=8.91, p<0.001, Locality $F_{7,103}$=2.34, p<0.001). All regions are significantly different from each other (Pairwise permutation MANOVAs, p<0.05), apart from West Andes and Amazon (Pairwise permutation MANOVAs, p=0.120). Both species show variation between geographic locations, with more variance explained by region in *H. erato* than *H. melpomene*. These results were confirmed by ManyGLM tests (Tables S13, S14, S15, S16).

Does genetic distance explain chemical divergence in H. erato *and* H. melpomene*?*

In *H. erato,* chemical distance is positively correlated with genetic distance, when accounting for geographic distance, although this correlation is weak for androconia (partial Mantel test, androconia, R=0.164, p=0.001; genitals, R=0.348, p=0.001). When we account for genetic distance, geographic distance is weakly correlated with androconial chemical distance and not correlated with genital chemical distance (partial Mantel test, androconia, R=0.151, p=0.002; genitals, R=-0.0775, p=0.966). These results were consistent when the analysis was repeated without unknown compounds (supplementary results).

Heliconius melpomene genitals show similar patterns to *H. erato*, but variation in the androconia is explained by geographic but not genetic distance. When accounting for geography, genetic divergence is not correlated with androconial chemical divergence, and only weakly with genital chemical divergence (partial Mantel test, androconia, R=0.02874, p=0.141, genitals, R=0.1203, p=0.001). When we first consider genetic distance, geographic distance is weakly positively correlated with androconial chemical distance, but not genital chemical distance (partial Mantel test, androconia, R=0.1795, p=0.002; genitals, R=-0.004, p=0.563).

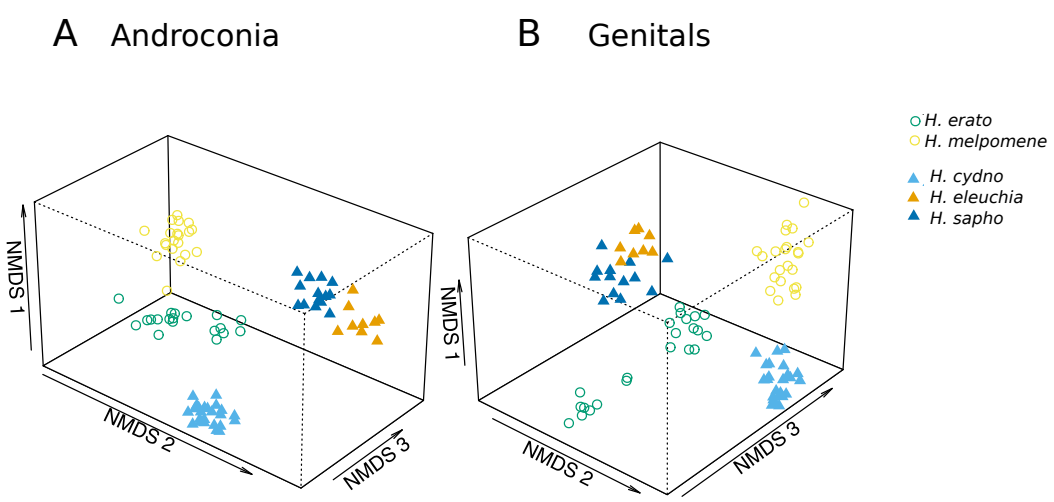


Figure 6. *NMDS (non-metric multidimensional scaling) plot illustrating in three dimensions the variation in chemical compounds of male* Heliconius *from Panama and Western Ecuador.* H. erato *and* H. melpomene *are co-mimics (circles), whilst* H. cydno*,* H. eleuchia *and* H. sapho *form a second co-mimicry group (triangles). A) Androconial chemical bouquets. Stress=0.098 B) Genital chemical bouquets. Stress=0.094.*

Is there evidence for similarity between co-mimics in chemical profile?

We investigated the effect of mimicry ring on chemical profile using individuals collected in Panama and western Ecuador from two mimicry rings (Fig. 6). Consistent with our interspecific analyses, we find that species is the main determinant of androconial and genital bouquets. *H. sapho* and *H. eleuchia* group closely in the NMDS visualisation, however, they are closely-related and so it is unclear whether this similarity is due to co-mimicry or shared ancestry. Especially for the androconia, *H. erato* and *H. melpomene* seem to be more similar than we might expect given their phylogenetic distance.

All the results described above show a consistent pattern when unidentified compounds were not included in the analysis (supplementary results). Interspecific analyses were also consistent when repeated without populations with a sample of fewer than 5 individuals (this removed 7 populations from androconial analysis and 5 from genital analysis) (supplementary results).

Discussion

Heliconius butterflies represent a continental-scale adaptive radiation (Kozak et al., 2015). Speciation in this group is often associated with divergence in wing colour pattern and pattern variation plays an important role in speciation and mate preference (Jiggins, 2008; Jiggins et al., 2001; Merrill et al., 2019, 2015, 2011; Sánchez et al., 2015). However, one of the surprising findings to emerge from comparative genomic analysis is the wealth of chemosensory genes (Heliconius Genome Consortium, 2012), suggesting that chemical signalling may play an important role in the biology of the system, such as host plant choice and mate choice. To begin to understand the role of chemical signalling in this radiation we have extensively surveyed both inter- and intraspecific variation of *Heliconius*' androconial and genital chemical profiles across the Neotropics. We find that most of the variation in chemical profile across our samples is explained by species, and we identify key chemicals serving as indicators for each species. Nonetheless, there is also intraspecific variation in chemical profiles. This variation is mainly quantitative in nature, with the exception of *H. erato cyrbia* which has compounds not found in other *H. erato* populations. We also find tentative evidence that co-mimics exhibit convergence in their

chemical profiles, supporting an earlier hypothesis (Mann et al., 2017). Our work sets the stage for further research into the biology and function of chemical profiles, and their role in within and between-species signalling.

It would be challenging to conduct behavioural experiments on large numbers of species and populations, and therefore identifying the behaviourally active components in pheromone blends across a radiation is beyond the scope of a single study. Other studies have also attempted to predict male sex pheromones without behavioural data, by selecting based on multiple criteria such as male-specificity and abundance (Bacquet et al., 2015). This step-wise selection of candidates focuses on within-species characteristics such as abundance, without considering the presence of the compound in other species. We hypothesised that consistent species-specific compounds are likely to be biologically important. We present an alternative method to detect candidate pheromones by evaluating both the presence of a compound across the geographic range of a species as well as the presence of the compound in other species. This approach has multiple advantages, including simple mathematics and the ability to evaluate combinations of compounds as well as single compounds. In *H. melpomene* the compounds identified as indicators for androconia and genitals, octadecanal and (*E*)-β-ocimene respectively, are both known to show electrophyisiological and behavioural activity (Byers et al., 2019; Schulz et al., 2008). Combining broad geographic sampling with indicator analysis therefore provides a promising approach to determine potential pheromone components in other species, which could be tested behaviourally. Our analyses have already identified a number of compounds that could now be tested functionally, such as the androconial compound geranylgeranylacetone in *H. erato.*

Chemical profiles are predicted to be highly species-specific if they are involved in species recognition during mating. For instance orchid bee chemical blends, presumably important for mating and species recognition, show high species-specificity, as well as within-species variability, which can be partly explained by geography (Brand et al., 2019; Weber et al., 2016; Zimmermann et al., 2006). We see similar patterns in *Heliconius*, with greater interspecific than intraspecific differences in chemical profiles. The magnitude of intraspecific differences is smaller in *Heliconius*, likely due to the fact that orchid bees collect their blends from the environment (Eltz et al., 1999). In both cases, species

identity is the best predictor of chemical divergence, with geographic location able to explain some intraspecific differences. One exception to this is *H. elevatus* which does not group separately from its co-mimic *H. melpomene* for genital compounds, despite the fact that these species are not especially closely related in the *Heliconius* phylogeny. Further samples are needed to confirm that this result is not due to the small sample of *H. elevatus* in this study. As in orchid bees, species differences are often consistent across a large geographic range, suggesting that they could be important for reproductive isolation between species (Weber et al., 2016).

We found a correlation between chemical distance and genetic distance. This suggests that neutral evolutionary forces are important in the evolution of chemical bouquets. The correlation between genital chemical distance and genetic distance is a much stronger correlation than previously reported (Estrada et al., 2011), possibly due to the quantitative nature of our data. The strong signal of neutrality suggests that the majority of compounds in the bouquets are neutrally evolving. For example, in the genital bouquet of *H. melpomene*, one compound, (*E*)-β-ocimene, can act by itself as an anti-aphrodisiac, with other components of the bouquet thought to moderate its evaporation rate (Schulz et al., 2008). In the future, focusing on the evolutionary patterns of only compounds which exhibit behavioural or electrophysiological responses, rather than the entire bouquet, may disentangle the processes involved in the evolution of these profiles.

Heliconius erato and *H. melpomene* both exhibit extensive colour pattern variation across their geographic range (Sheppard et al., 1985) and these populations also differ in their androconial and genital bouquets. Whilst traditionally predicted to be under stabilising selection, intraspecific variation between populations in chemical profiles has been documented in other Lepidoptera (Carde and Allison, 2016). Chemical divergence in putative male sex pheromones between populations of *Bicyclus anynana* is reported to be as large as differences between *Bicyclus* species, and is greater than predicted by genetic divergence (Bacquet et al., 2016). This is in contrast to what we find here, where interspecific differences are much greater than intraspecific ones.

Interestingly, *Heliconius erato cyrbia* produces many unique genital compounds and is also the most genetically divergent *H. erato* population in our study, suggesting

that genetic drift is important for the evolution of chemical profiles in *Heliconius*. Across all *H. erato* populations we find a correlation between chemical distance and genetic distance, which is weaker for androconial bouquets. In *H. melpomene*, genetic distance is also weakly correlated with genital chemical divergence. These correlations suggest that some of the geographic variation between populations could be neutral, with stochastic processes important for bouquet evolution in *Heliconius*. In contrast, androconial chemical variation in *H. melpomene* is better explained by geographic distance. This might imply that other evolutionary forces are important for chemical profile evolution in *H. melpomene*.

One factor potentially involved in geographic variation is larval host plant use. Feeding on different host plants as larvae affects the production of some minor components of both androconial and genital chemical bouquets (Darragh et al., 2019a). The major components, however, are unaffected by larval host plant, suggesting that any dietary precursors required for compound production are present in different *Passiflora* species (Darragh et al., 2019a). In Panama, *H. cydno* and *H. melpomene* both feed on *P. menispermifolia* (Merrill et al., 2013), and yet have different chemical profiles, highlighting that from the same precursors different species can produce different compounds. Furthermore, it is often unclear which is the major *Passiflora* host plant of any particular *Heliconius* population. The distribution of *Passiflora* species varies geographically (Benson, 1978; Benson et al., 1975), and both host preference and level of host-specificity vary between populations of the same *Heliconius* species (Castro et al., 2018). A greater understanding of the variation in larval diet of *Heliconius* across the Neotropics will help us understand how much geographic variation in chemical profile can be attributed to host plant use.

Heliconius butterflies are an excellent example of visual mimicry, with different species converging on the same warning colour patterns (Merrill et al., 2015; Sheppard et al., 1985; Sherratt, 2008). It has been suggested that chemical compounds could also contribute to mimicry between species (Dettner and Liepert, 1994; Mann et al., 2017). In this study, we find tentative evidence that individuals within particular co-mimicry groups, such as *H. melpomene* and *H. erato*, have more similar chemical profiles than expected. Most known examples of chemical mimicry come from systems of deception,

for example, mimicry of ant alarm pheromones by rove beetles to avoid predation, rather than mimicry of aposematic warning signals (Dettner and Liepert, 1994; Stoeffler et al., 2007; Vereecken and McNeil, 2010). We suggest that in *Heliconius* different components of the bouquet could be important for chemical mimicry and species recognition, reducing conflict between these selection pressures.

Convergence of genital bouquets between co-mimics could be due to the anti-aphrodisiac function of these compounds (Gilbert, 1976; Schulz et al., 2008). Anti-aphrodisiac compounds are transferred from males to females during mating to deter future matings from other males. Convergence in wing pattern between co-mimics could result in harassment not only by conspecific but also heterospecific males (Estrada and Jiggins, 2008). The use of the same anti-aphrodisiac by co-mimics could combat interspecific attraction by deterring males of both species, as highlighted by the production of (*E*)-β-ocimene by *H. erato* and *H. melpomene*, as well as other *Heliconius* species (Estrada et al., 2011).

Compounds could also play a role in predator deterrence. Genital compounds were originally suggested to form part of the anti-predation signal (Eltringham, 1925). We detected 2-*sec*-butyl-3-methoxypyrazine in the genitals of *H. melpomene*, *H. cydno* and *H. timareta*, and 2-isobutyl-3-methoxypyrazine in the genitals of *H. melpomene* and *H. cydno*, both compounds known to deter predators in the wood tiger moth (Burdfield-Steel et al., 2018; Rojas et al., 2019, 2018, 2017). More generally, methoxypyrazines act as warning odours in other insects (e.g. Lepidoptera, Rothschild, Moore & Brown, 1984; fireflies, Vencl et al., 2016), effective against avian predators (Guilford et al., 1987). Further investigation will be required to determine if odours of *Heliconius* butterflies act as anti-predation signals.

Overall, our study reveals strong species differences in bouquets and the presence of species-specific compounds, as well as intraspecific variation. A pattern of species-specificity alongside intraspecific variation could be the result of a balance between stabilising selection towards a species stereotype, sexual selection promoting diversity, and geographic segregation alongside selection and drift. A challenge for the field is the feasibility of testing for the biological relevance of hundreds of compounds in many

species, but we hope that our innovative analysis will stimulate not only further targeted functional studies of putatively important compounds, but also large chemical profile surveys in other study systems of evolutionary interest.

Supplementary Information

Does removing unidentified compounds or poorly sampled populations affect models of interspecific variation in chemical profiles?

We repeated the interspecific analysis without unidentified compounds. When repeated without unidentified compounds, species still significantly differ in their androconial bouquet, with species identity accounting for 60% of the overall variation in chemical profiles (PERMANOVA, Species, $F_{6,251}$=77.33, p<0.001). A further 4% of variation can be explained by region (Amazon/Eastern Andes/Western Andes/Panama), and 3% by locality nested within region (PERMANOVA, Region, $F_{3,251}$=9.96, p<0.001; (Region/Locality), $F_{8,251}$=2.49, p<0.001). Finally, 4% of variation is explained by an interaction between species and region (PERMANOVA, Species*Region $F_{6,251}$=4.77, p<0.001;). Results were also similar for genital bouquets, with species identity still explaining the highest amount of variation, accounting for 44% of the variation in chemical profiles (PERMANOVA, Species, $F_{6,274}$=45.44, p<0.001). A further 6% of variation can be explained by region (Amazon/Eastern Andes/Western Andes/Panama), and 4% by locality nested within region (PERMANOVA, Region, $F_{3,274}$=12.74, p<0.001; (Region/Locality), $F_{8,274}$=2.85, p<0.001). Finally, 6% of variation is explained by an interaction between species and region (PERMANOVA, Species*Region $F_{6,274}$=6.32, p<0.001).

We also repeated the interspecific analyses removing populations with fewer than 5 individuals. Again, similar to removing unidentified compounds, and the full dataset, species identity accounts for 58% of overall variation in chemical profiles, with a further 5% explained by region, 3% by locality nested within region, and 4% by an interaction between species and region (PERMANOVA, Species, $F_{5,227}$=78.26, p<0.001; Region, $F_{3,227}$=10.19, p<0.001; (Region/Locality), $F_{7,227}$=2.98, p<0.001; Species*Region $F_{4,227}$=6.92, p<0.001;). Results were also consistent for genital bouquets, with species identity still explaining the highest amount of variation, accounting for 51% of the variation in chemical profiles, with a further 5% explained by region, 4% by locality nested within region, and 5% by an interaction between species and region (PERMANOVA, Species,

$F_{5,255}$=69.49, p<0.001; Region, $F_{3,255}$=12.37, p<0.001; (Region/Locality), $F_{8,255}$=3.05, p<0.001; Species*Region $F_{4,255}$=9.13, p<0.001).

Does removing unidentified compounds or poorly sampled populations affect correlations between divergence in chemical profile with genetic and geographic distance?

Correlations with genetic and geographic distances were also consistent with results including all compounds. When controlling for geographic distance, genetic divergence is strongly correlated with both androconial and genital chemical divergence (Mantel test, androconia, r=0.7897, p=0.001; genitals, r=0.5203, p=0.001). When controlling for genetic distance, geographic distance is significantly but weakly correlated with chemical divergence (Mantel test, androconia, r=0.06739, p=0.002; genitals, r=0.059, p=0.003).

Removing populations with fewer than 5 individuals also gave consistent results. When controlling for geographic distance, genetic and chemical divergence remain strongly correlated (Mantel test, androconia, r=0.7978, p=0.001; genitals, r=0.71, p=0.001). Again, when controlling for genetic distance, geographic distance is significantly but weakly correlated with androconial and genital chemical divergence (partial Mantel test, androconia, r=0.082, p=0.001; genitals, r=0.0439, p=0.006).

Does removing unidentified compounds affect models of intraspecific variation in chemical profiles?

Individuals of *H. erato* still strongly group by region when unidentified compounds are removed from the analysis, with 27% of variation in androconial profiles being explained by region and 11% by locality nested within region (PERMANOVA, Region $F_{3,87}$=11.49, p<0.001, Locality $F_{6,87}$=2.30, p<0.001). Again, this is similar for *H. erato* genital compounds. Region explains 35% of variation, and 8% is explained by locality nested within region (PERMANOVA, Region $F_{3,91}$=16.76, p<0.001, Locality $F_{6,91}$=1.98, p<0.01).

Again, we found consistent results without unidentified compounds for *H. melpomene.* The same amount of variation was explained in models with and without

unidentified compounds included, with region explaining 18% of variation in androconial compounds (PERMANOVA, Region $F_{3,86}$=6.05, p<0.01). For *H. melpomene* genital compounds, 20% of variation is explained by region, 12% by locality nested within region, as in models with all compounds included (PERMANOVA, Region $F_{3,103}$=9.05, p<0.001, Locality $F_{7,103}$=2.34, p<0.001).

Does removing unidentified compounds affect correlations between intraspecific chemical divergence with genetic and geographic distance?

Results were again very similar without unidentified compounds included in the analysis. In *H. erato* both androconial and genital chemical distance are positively correlated with genetic distance, even when accounting for geographic distance (partial Mantel test, androconia, R=0.148, p=0.001; genitals, R=0.280, p=0.001). When unidentified compounds are removed, we still find a weak positive correlation between geographic distance and androconial, but not genital distance, accounting for genetic distance (partial Mantel test, androconia, R=0.155, p=0.002; genitals, R=-0.0171, p=0.656).

For *H. melpomene*, genital bouquet divergence, but not androconial bouquet divergence, is correlated with genetic distance when accounting for geography (partial Mantel test, androconia, R=0.02602, p=0.169, genitals, R=0.112, p=0.001). When we first consider genetic distance, geographic distance is not positively correlated with genital chemical distance, however, it is positively correlated with androconial chemical distance (partial Mantel test, androconia, R=0.1729, p=0.002; genitals, R=0.003, p=0.439). These results are consistent with tests including all compounds.

Comparison of within and between spcies and locality Bray-Curtis distances

We compared within and between species and locality Bray-Curtis distances. We focused on *H. erato* and *H. melpomene*, as these species were collected in the most localities. We calculated a Bray-Curtis distance matrix and then used the function "dist_groups" from the package usedist to calculate distances between individuals of

different groups (Bittinger, 2017). We add statistical comparisons to the violin plots using the function “stat_compare_means” from the package ggpubr (Kassambara, 2019). For both androconia and genitals, the mean chemical distance between individuals is greater between species (androconia, 0.971; genitals, 0.915) than within species (androconia, 0.554; genitals, 0.573). The mean chemical distance between individuals is also greater between localities (androconia, 0.564; genitals, 0.584) than within localities (androconia, 0.457; genitals, 0.466). However, the magnitude of this difference is much smaller, confirming the PERMANOVA and ManyGLM analyses that most variation is explained by species and not geographic location.

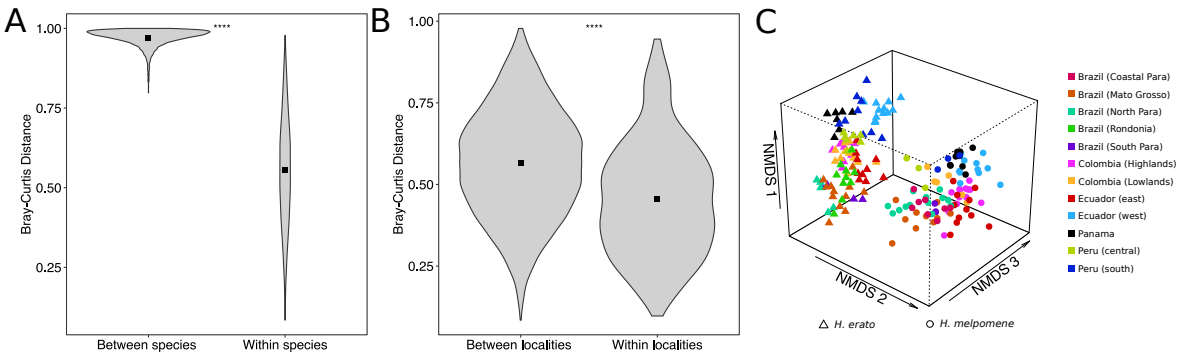


Figure S1. *Pairwise androconial Bray-Curtis distances between individuals of* H. erato *and* H. melpomene *A) between and within species and B) between and within localities of the same species. C) NMDS (non-metric multidimensional scaling) plot illustrating the variation in androconial chemical compounds of male* H. erato *and* H. melpomene *from different localities.*

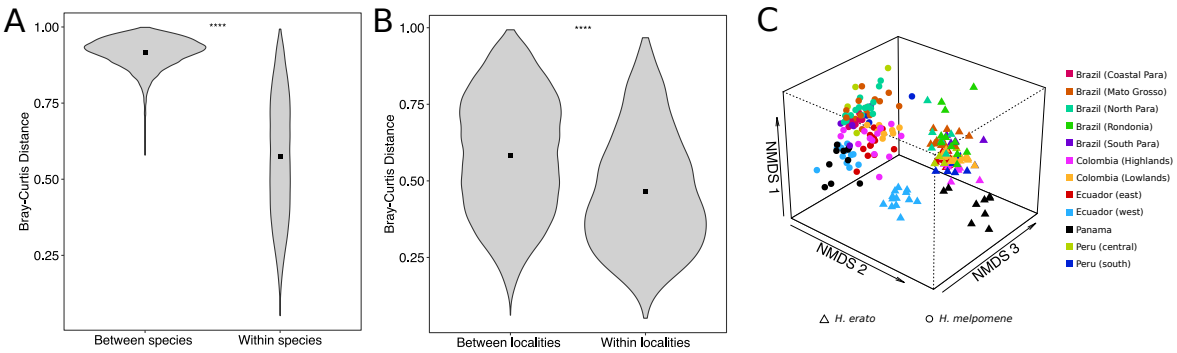


Figure S2. *Pairwise genital Bray-Curtis distances between individuals of* H. erato *and* H. melpomene *A) between and within species and B) between and within localities of the same species. C) NMDS (non-metric multidimensional scaling) plot illustrating the variation in genital chemical compounds of male* H. erato *and* H. melpomene *from different localities.*

Table S1: *Number of androconial and genital samples collected for each locality*

Locality	Species	No. androconial samples	No. genital samples
Panama	*H. cydno*	12	12
Panama	*H. erato*	7	9
Panama	*H. melpomene*	9	10
Panama	*H. sapho*	12	11
Colombia (highlands)	*H. erato*	10	9
Colombia (highlands)	*H. melpomene*	13	14
Colombia (highlands)	*H. timareta*	7	8
Colombia (lowlands)	*H. erato*	12	13
Colombia (lowlands)	*H. melpomene*	7	8
Ecuador (east)	*H. elevatus*	2	2
Ecuador (east)	*H. erato*	10	11
Ecuador (east)	*H. melpomene*	11	13
Ecuador (east)	*H. timareta*	2	3
Ecuador (west)	*H. cydno*	12	12
Ecuador (west)	*H. eleuchia*	9	8
Ecuador (west)	*H. erato*	12	13
Ecuador (west)	*H. melpomene*	11	12
Ecuador (west)	*H. sapho*	2	2
Peru (central)	*H. erato*	7	6
Peru (central)	*H. melpomene*	3	4
Peru (south)	*H. erato*	9	9
Peru (south)	*H. melpomene*	3	3
Brazil (Coastal Para)	*H. melpomene*	8	8
Brazil (Mato Grosso)	*H. erato*	12	12
Brazil (Mato Grosso)	*H. melpomene*	10	15
Brazil (North Para)	*H. erato*	5	6
Brazil (North Para)	*H. melpomene*	13	15
Brazil (South Para)	*H. elevatus*	2	2
Brazil (South Para)	*H. erato*	4	5
Brazil (South Para)	*H. melpomene*	3	6
Brazil (Rondonia)	*H. elevatus*	3	3
Brazil (Rondonia)	*H. erato*	10	11

Table S2: *Genome samples of* H. erato *and* H. melpomene *races including number of androconial (A) and genital (G) samples included in analysis for each race. All samples were males. Newly sequenced individuals are denoted with a star next to their ID. More information on individuals can be found on the public database https://heliconius.ecdb.io/*

Locality	Taxon name	A	G	ID	Lat.	Lon.
Colombia (highlands)	*H. m. bellula*	13	14	CAM040049*	1.217	-76.683
Colombia (lowlands)	*H. m. malleti*	7	8	CS002311	1.814	-75.669
Ecuador (east)	*H. m. malleti*	11	13	CAM016540	-1.061	-77.668
Ecuador (west)	*H. m. cythera*	11	12	14N015*	0.185	-78.853
Brazil (Mato Grosso)	*H. m. burchelli*	3	5	SR281*	-13.814	-56.404
Brazil (Coastal Para)	*H. m. intersectus*	1	1	KK291*	-1.070	-46.745
Brazil (South Para)	*H. m. madeira*	2	5	SR391*	-4.066	-54.847
Brazil (North Para)	*H. m. melpomene*	13	15	SR178*	-1.937	-54.626
Brazil (Mato Grosso)	*H. m. penelope*	4	7	SR358*	-13.691	-57.706
Brazil (Coastal Para)	*H. m. thelxiope*	7	7	KK288*	-1.070	-46.745
Peru (south)	*H. m. schunkei*	3	3	KK544*	-13.204	-70.768
Peru (central)	*H. m. xenoclea*	3	4	KK309*	-11.0354	-75.407
Panama	*H. m. rosina*	9	10	CAM001841	9.076	-79.659
Colombia (highlands)	*H. e. dignus*	10	9	CAM040113*	1.214	-76.690
Colombia (lowlands)	*H. e. lativitta*	12	13	CAM040160*	0.956	-76.409
Ecuador (west)	*H. e. cyrbia*	12	13	CAM040545*	0.151	-78.770
Ecuador (east)	*H. e. lativitta*	10	11	CAM041030*	-1.059	-77.702
Brazil (South Para)	*H. e. amazona*	4	5	SR122*	-4.066	-54.847

Brazil (Mato Grosso)	*H. e. phyllis*	10	8	SR230*	-10.891	-55.440
Brazil (Rondonia)	*H. e. venustus*	9	10	SR314*	-12.806	-60.297
Peru (south)	*H. e. amphitrite*	9	9	KK464*	-12.955	-72.656
Peru (central)	*H. e. emma*	1	1	KK402*	-10.298	-74.935
Peru (central)	*H. e. microclea*	4	4	KK338*	-11.055	-75.419
Panama	*H. e. demophoon*	7	9	Pet_ED3	-9.129	79.715

Table S3. *Model selection table for PERMANOVA models based on AIC scores. If two models were within two AIC points of each other, we chose the simpler model as the most parsimonious. Best fit models are highlighted in bold.*

Model	**Residual Sum of Squares**	**DF**	**AIC**
Interspecific androconia			
Chemical profile ~ Species + Region + (Region/Locality) + Species*Region + Species*(Region/Locality)	26.436	31	889.2356
Chemical profile ~ Species + Region + (Region/Locality) + Species*Region	28.185	23	889.3774
Chemical profile ~ Species + Region + (Region/Locality)	31.760	17	907.4649
Chemical profile ~ Species + Region	34.377	9	911.4244
Chemical profile ~ Species	38.07	6	931.1351
Interspecific genitals			
Chemical profile ~ Species + Region + (Region/Locality) + Species*Region + Species*(Region/Locality)	31.887	31	1016.104
Chemical profile ~ Species + Region + (Region/Locality) + Species*Region	**33.748**	**23**	**1015.703**
Chemical profile ~ Species + Region + (Region/Locality)	39.008	17	1043.535
Chemical profile ~ Species + Region	42.151	9	1048.847
Chemical profile ~ Species	47.165	6	1073.751
H. erato androconia			
Chemical profile ~ Region + (Region/Locality)	8.79	9	211.2867
Chemical profile ~ Region	10.3793	3	213.9033
H. erato genitals			
Chemical profile ~ Region + (Region/Locality)	9.1105	9	223.2669
Chemical profile ~ Region	10.9982	3	228.5917
H. melpomene androconia			
Chemical profile ~ Region + (Region/Locality)	9.8523	11	223.0307
Chemical profile ~ Region	11.7698	3	222.502
H. melpomene genitals			
Chemical profile ~ Region + (Region/Locality)	12.3471	10	283.3958
Chemical profile ~ Region	14.5178	3	286.2391

Table S4. *Analysis of deviance model selection table for multivariate generalised linear models based on likelihood ratio tests. Each model is compared to the model above it. Best fit models are highlighted in bold.*

Model	**ΔDeviance**	**Residual DF**	**p-value**
Interspecific androconia			
Chemical profile ~ Species + Region + (Region/Locality) + Species*Region + Species*(Region/Locality)		220	
Chemical profile ~ Species + Region + (Region/Locality) + Species*Region	-740.7	228	0.001
Interspecific genitals			
Chemical profile ~ Species + Region + (Region/Locality) + Species*Region + Species*(Region/Locality)		243	
Chemical profile ~ Species + Region + (Region/Locality) + Species*Region	-1854	251	0.063
Chemical profile ~ Species + Region + (Region/Locality)	-2787	257	0.134
Chemical profile ~ Species + Region	-6481	265	0.026
H. erato androconia			
Chemical profile ~ Region + (Region/Locality)		78	
Chemical profile ~ Region	-1091	84	0.001
H. erato genitals			
Chemical profile ~ Region + (Region/Locality)		82	
Chemical profile ~ Region	-2721	88	0.001
H. melpomene androconia			
Chemical profile ~ Region + (Region/Locality)		75	
Chemical profile ~ Region	-1582	83	0.001
H. melpomene genitals			
Chemical profile ~ Region + (Region/Locality)		93	
Chemical profile ~ Region	-1777	100	0.001

Table S5: *Androconial compounds identified in at least half of all individuals from any population. The Kovats gas chromatographic retention index (RI) is reported for each compound. Mean amounts (ng) in* H. cydno *(CYD),* H. eleuchia *(ELEU),* H. elevatus *(ELEV),* H. erato *(ERAT),* H. melpomene *(MEL),* H. sapho *(SAPH),* H. timareta *(TIM).*

Compound	RI	CYD	ELEU	ELEV	ERAT	MEL	SAPH	TIM
Unknown	1018	0.00	0.00	0.00	0.63	0.37	0.00	0.00
(Z)-β-Ocimene	1030	0.00	0.00	123.97	0.24	40.92	0.00	0.00
Phenylacetaldehyde	1036	1.03	0.26	0.00	0.07	0.02	0.00	1.30
Acetophenone	1065	0.00	0.00	0.00	0.06	0.00	0.00	0.00
cis-Linalool oxide	1082	0.88	0.00	3.71	0.00	0.00	0.00	0.00
o-Guaiacol	1090	0.08	0.00	0.00	0.00	0.39	0.03	0.22
Nonanal	1105	3.45	0.75	3.91	2.87	21.05	1.02	15.41
3E,5E-2,6-Dimethyl-1,3,5,7-octatraen	1120	0.00	0.00	0.59	0.00	0.23	0.00	0.00
Phenylacetonitrile	1139	6.88	0.00	0.86	0.08	0.00	0.00	8.49
Unknown	1145	0.00	0.00	0.00	0.89	0.00	0.00	0.00
Benzyl acetate	1165	0.00	0.00	0.00	0.61	0.19	0.10	0.00
Napthalene	1175	6.04	1.20	8.29	3.45	4.66	1.87	0.38
Unknown	1184	0.00	0.00	0.00	3.19	0.00	0.00	0.00
Unknown	1188	0.04	0.00	0.00	8.80	1.17	0.32	0.00
Methyl Salicylate	1189	9.72	2.99	12.43	2.62	2.82	1.58	0.74
Decanal	1198	0.12	0.00	0.00	0.00	0.84	0.16	0.00
4-Vinylphenol	1215	0.00	13.98	0.00	0.00	0.00	2.71	0.00
Benzothiazole	1215	0.00	0.00	0.00	0.01	0.13	0.00	0.00
Unknown	1236	0.00	0.00	0.00	0.05	0.03	0.00	7.28
Unknown	1268	1.47	0.00	0.00	0.01	0.54	0.00	0.00
Nonanoic acid	1269	0.00	0.00	0.24	0.02	1.24	0.00	0.00
3-Undecanone	1275	0.00	0.00	0.00	13.79	0.00	0.00	0.08
1-Tridecene	1290	0.00	0.00	0.00	4.26	0.00	0.00	0.00
Indole	1295	0.00	0.00	0.04	0.00	0.19	0.00	0.00
Unknown aromatic	1299	0.00	0.00	0.72	0.43	0.00	0.00	0.15
Unknown	1353	0.00	0.00	0.00	0.06	0.18	0.00	1.52
Glycerol ester	1356	0.20	0.00	0.00	0.87	0.53	0.47	102.62
Unknown	1366	0.04	0.00	0.00	0.22	0.14	0.00	0.68
1,3-Dimethyluracil	1371	0.00	0.00	0.37	0.40	0.05	0.03	0.00
Unknown	1371	0.13	0.00	0.00	0.06	0.02	0.00	4.47
Unknown alkene	1382	0.00	0.00	0.00	0.97	0.00	0.00	0.00
Unknown ester	1392	0.00	8.58	0.00	0.00	0.00	7.27	0.00
Butanoic acid ester	1394	0.00	0.16	0.00	1.16	0.00	0.60	0.00
Unknown	1396	0.06	0.00	0.00	10.72	0.07	0.00	0.00
Unknown	1424	0.00	0.00	0.00	1.10	0.00	0.00	0.00

2-Methoxy-4-propylphenol	1436	0.00	0.00	24.40	0.00	0.02	0.00	0.00
Unknown	1444	0.44	0.00	0.00	1.79	0.04	0.00	0.88
Unknown	1445	0.10	0.00	0.00	0.00	0.68	0.69	0.08
Unknown	1483	0.00	0.00	0.00	0.00	0.00	0.00	1.34
5-Decanolide	1493	0.00	0.00	0.00	0.00	0.00	0.00	22.31
Methyl 4-hydroxy- methoxybenzoate	1517	0.00	0.00	0.00	1.35	0.12	66.11	0.00
Unknown aromatic	1519	0.00	0.00	0.00	0.00	2.87	0.00	0.00
Unknown aromatic	1519	0.07	0.00	0.00	0.00	2.98	0.00	0.00
d-Cadinene	1522	0.00	0.00	0.00	1.81	0.06	0.00	0.00
Homovanillyl alcohol	1523	0.04	0.00	695.61	0.74	4.02	0.00	3.23
Dihydroactinidiolide	1532	8.57	0.48	0.00	44.91	22.69	5.88	28.55
Calacorene	1540	0.00	0.00	0.21	0.04	0.04	0.00	0.00
Mellein	1540	0.00	44.75	89.60	179.38	27.21	14.19	0.00
Unknown aromatic	1563	0.73	0.00	0.31	0.00	0.20	0.08	0.00
Unknown	1571	0.00	0.00	0.00	0.00	0.33	0.00	0.00
Unknown alkene	1581	0.00	0.00	0.00	0.68	0.00	0.00	0.00
Methyl 3,4- dimethoxybenzoate	1582	0.00	0.00	0.00	0.00	4.74	0.00	0.00
Unknown	1638	0.00	0.00	0.00	0.01	1.18	0.00	2.40
Unknown	1639	0.00	0.00	1.18	0.76	0.42	0.15	1.83
d-Cadinol	1644	0.34	0.00	0.00	157.03	0.09	0.00	0.00
Unknown	1649	0.00	0.00	0.00	0.00	0.25	0.00	0.00
Syringaldehyde	1661	937.77	0.00	574.45	0.30	488.42	183.68	1043.19
Heptadecene	1679	0.00	0.00	0.46	0.57	0.23	0.00	0.00
Unknown	1704	0.00	0.00	0.00	2.30	0.00	0.00	0.00
3,5-Dimethoxy-4- hydroxybenzyl alcohol	1707	4.93	0.00	1.91	0.01	14.08	9.31	52.82
Unknown	1714	0.00	0.00	0.00	3.97	0.00	0.00	0.00
1-(3,5-Dimethoxy-4-hydroxybenzyl) ethanone	1735	0.79	0.00	1.59	0.00	1.42	0.00	0.54
Unknown aromatic	1738	0.00	0.00	0.00	0.00	0.85	0.00	0.00
Unknown hydrocarbon	1750	1.62	1.12	3.38	3.86	2.60	2.59	4.37
1*H*-Indol-3-ethanol	1754	0.00	0.00	0.21	0.00	0.90	0.00	0.00
Unknown aromatic	1757	0.00	0.00	0.00	0.00	0.45	0.00	0.00
Ethyl benzoate	1762	0.42	0.00	1.69	0.90	0.45	1.38	0.73
Benzyl benzoate	1766	0.00	0.00	0.00	0.00	0.32	0.10	0.00
Unknown	1766	0.00	0.13	0.00	0.00	0.52	0.00	0.00
Unknown	1771	0.00	0.00	0.70	0.17	0.13	0.03	0.00
Unknown	1780	0.00	0.00	0.00	0.25	0.00	0.00	0.00
Unknown aromatic	1797	0.00	0.00	0.00	0.00	0.21	0.06	2.87
Octadecane	1800	0.02	0.00	2.94	1.06	0.80	0.03	0.00
Hexadecadien-15-olide	1807	0.00	0.00	0.00	54.93	0.00	0.00	0.00
Unknown aromatic sulfur	1807	0.65	0.00	0.20	0.01	2.05	0.00	0.10
Hexadecanal	1815	0.00	0.00	0.00	2.32	0.55	0.00	0.00
Methyl 1*H*-indol-3-acetate	1822	0.00	0.00	0.00	0.00	1.44	0.00	0.34

Neophytadiene	1836	0.00	399.37	0.00	0.00	0.00	24.48	0.00
Unknown	1841	0.13	0.00	0.00	0.04	0.00	0.00	2.93
Hexahydrofarnesylacetone	1843	0.00	813.34	0.00	0.00	0.00	177.49	0.00
Methyl 1*H*-indol-3-carboxylate	1853	0.00	0.00	5.54	0.03	1.49	0.00	1.56
Neophytadiene	1861	0.00	102.95	0.00	0.00	0.00	7.06	0.00
Benzyl salicylate	1870	0.23	20.44	5.48	0.18	4.92	51.30	0.00
Unknown branched alcohol	1873	0.00	0.00	0.00	0.00	1.29	0.00	0.00
Nonadecene	1877	0.00	0.00	0.51	0.52	0.21	0.00	0.00
Neophytadiene	1878	0.00	163.30	0.00	0.00	0.00	8.71	0.00
1-Hexadecanol	1884	0.00	0.00	1.15	479.25	2.53	0.00	0.00
Nonadecane	1899	7.34	0.00	11.06	0.43	0.23	0.00	0.00
Unknown	1915	0.05	0.00	3.35	1.73	1.11	0.99	0.00
Heptadecanal	1918	0.00	1.61	0.00	0.00	3.70	0.00	0.00
Hexadecanoic acid	1960	1.28	0.49	52.75	0.81	0.76	0.00	0.00
Unknown hydrocarbon	1962	0.00	0.00	0.54	0.58	0.24	0.71	0.00
Unknown	1962	0.21	0.26	0.55	0.73	0.16	0.58	0.76
Unknown	1970	2.95	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	1979	0.00	0.00	0.00	0.79	2.03	0.00	0.00
(Z)-9-Octadecenal	1990	0.51	0.00	0.00	0.00	143.53	0.00	0.00
Hexadecylacetate	1991	0.00	0.00	0.00	4.79	0.00	0.00	0.00
Icosane	2000	15.52	0.00	37.08	0.09	0.20	0.00	0.00
Octadecanal	2013	0.68	0.00	0.00	0.34	3703.41	12.65	5.88
Isopropyl Palmitate	2020	6.55	2.28	0.00	0.11	0.00	0.00	0.00
Unknown	2036	1.12	0.00	0.00	0.00	0.00	0.00	0.00
9-Octadecen-1-ol	2042	0.00	0.00	0.00	0.00	38.95	0.00	0.00
Unknown macrolide	2045	0.00	0.00	0.00	1.17	0.00	0.00	0.00
Unknown	2047	1.13	0.00	0.00	0.00	0.00	0.00	0.00
Nonadecanal (methyl, branched)	2054	0.00	0.00	0.00	0.00	22.87	0.00	0.00
Ethyl 4-hydroxy-3,5- dimethoxybenzoate	2057	1.54	0.00	3.48	0.10	12.00	0.00	1.37
Methyloctadecanal	2064	0.00	0.00	0.00	0.00	45.62	0.00	0.00
Henicosene	2068	1.02	0.00	2.97	0.95	0.00	0.00	0.00
Henicosene	2072	0.55	0.00	4.00	1.65	0.00	0.00	0.00
Methyloctadecanal	2072	0.00	0.00	0.00	0.00	20.58	0.00	0.00
Methyloctadecanal	2077	0.00	0.00	0.00	0.00	152.20	0.00	0.00
1-Octadecanol	2083	0.00	0.00	2.20	39.87	1189.47	4.59	0.42
Henicosene	2086	0.40	0.00	8.14	0.86	0.10	0.00	0.00
(*Z*)-16-Methyl-9-octadecenol	2092	0.00	0.00	0.00	0.00	31.95	0.00	0.00
Henicosane	2100	2543.56	0.39	4554.83	0.71	181.97	1.02	141.39
Unknown	2110	12.18	0.00	0.00	0.00	0.00	0.00	0.00
(*E*)-Phytol	2112	0.00	1071.10	0.00	0.00	0.00	680.94	0.00
Unknown	2112	0.00	0.00	0.00	0.00	15.88	0.65	0.00

Unknown alkene/alcohol	2127	0.00	0.00	0.00	0.00	68.88	0.00	0.00
Unknown aromatic	2130	2.83	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	2133	0.00	0.00	0.00	0.00	97.03	0.00	0.00
Unknown amide	2134	0.12	0.00	0.74	0.13	0.10	0.00	0.00
Methyloctadecan-1-ol	2138	0.00	0.00	0.00	0.00	32.53	0.00	0.00
Phytenal	2141	0.00	5.17	0.00	0.00	0.00	4.56	0.00
Unknown amide	2157	0.12	0.21	1.28	0.32	0.39	0.00	0.00
Unknown	2160	0.00	0.00	0.46	0.14	0.17	0.00	0.00
Ethyl oleate	2160	0.00	0.00	0.00	0.00	3.01	0.00	0.00
Unknown	2165	0.75	30.24	0.00	0.05	0.00	0.00	0.00
Unknown	2180	0.19	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	2189	0.00	0.00	0.32	0.38	0.23	0.00	0.00
(*Z*)-11-Icosenal	2192	4.47	0.00	0.38	0.00	311.60	0.00	0.00
Docosane	2200	12.47	0.00	48.09	1.85	2.83	0.20	0.58
Unknown	2215	0.00	0.00	2.72	0.00	1.85	0.22	0.00
Phytol	2217	0.00	10.72	0.00	0.00	0.00	1.98	0.00
Icosenol	2255	20.19	0.00	7.95	0.00	174.52	0.11	6.54
17-Methyloctadecylacetate	2271	0.00	0.00	0.00	0.00	2.71	0.00	0.00
Tricosene	2276	1.56	0.00	3.66	0.39	18.96	0.00	0.00
Unknown	2277	0.00	0.00	0.38	0.61	0.72	0.00	0.00
Icosanal	2283	0.00	0.00	0.00	0.01	118.07	6.57	0.00
Tricosane	2300	73.71	1.46	243.92	5.94	13.71	0.15	1.27
Unknown	2310	0.00	0.00	0.00	19.66	0.00	0.00	0.00
11-Icosenol	2314	0.00	0.00	0.00	0.00	1.86	0.00	0.00
Unknown fatty acid amide	2325	0.98	0.89	1.20	1.04	0.57	1.35	0.70
Unknown	2346	0.00	0.19	0.14	0.12	0.24	0.30	0.00
Unknown amide	2347	4.61	0.00	60.92	28.34	29.68	0.00	0.00
Unknown terpene	2351	0.00	0.00	0.00	28.82	0.00	0.00	0.00
Unknown	2353	5.58	4.99	7.90	4.72	8.51	3.96	6.57
Unknown	2354	0.00	0.00	0.00	1.68	0.00	0.00	0.00
Unknown amide	2363	0.00	0.00	0.18	0.17	0.24	0.00	0.00
Geranylgeranylacetone	2382	0.00	0.00	0.00	304.07	0.00	0.00	0.00
(*Z*)-13-Docosenal	2396	0.39	0.00	0.00	0.00	87.58	0.00	0.00
Tetracosane	2400	0.30	0.32	13.32	12.99	5.22	0.00	7.02
19-Methylicosylacetate	2422	0.00	37.61	0.00	3.20	4.25	2.37	0.00
(*Z*)-13-Docosen-1-ol	2464	0.00	0.00	0.00	0.88	18.48	1.68	0.00
Unknown	2490	0.00	0.00	0.58	0.39	0.03	0.00	0.00
Pentacosane	2500	4.39	16.62	44.57	30.48	13.06	4.49	12.00
11-Methylpentacosane	2532	0.00	2.12	70.91	0.00	0.88	0.18	0.00

Table S6: *Genital compounds identified in at least half of all individuals from any population. The Kovats gas chromatographic retention index (RI) is reported for each compound. Mean amounts (ng) in* H. cydno *(CYD),* H. eleuchia *(ELEU),* H. elevatus *(ELEV),* H. erato *(ERAT),* H. melpomene *(MEL),* H. sapho *(SAPH),* H. timareta *(TIM).*

Compound	**RI**	**CYD**	**ELEU**	**ELEV**	**ERAT**	**MEL**	**SAPH**	**TIM**
Benzaldehyde	961	2.30	0.00	0.00	0.54	0.00	0.00	4.33
β-Myrcene	990	0.00	0.00	5.96	0.00	9.32	0.00	0.00
Unknown	1018	0.00	0.00	0.43	0.33	0.38	589.07	0.00
Limonene	1024	0.00	0.00	0.00	3.95	0.02	0.00	0.00
(*Z*)-β-Ocimene	1030	5.21	21.55	30.71	54.25	47.61	13.01	0.00
Phenylacetaldehyde	1036	0.81	0.00	0.00	0.85	0.00	0.00	0.26
(*E*)-β-Ocimene	1054	4.30	232.69	16663.34	2556.24	22841.08	26.96	47.32
Unknown	1076	0.00	0.00	0.00	145.11	0.00	0.00	0.00
o-Guaiacol	1090	1.95	0.13	0.00	0.13	0.15	0.00	0.00
Nonanal	1099	0.00	0.00	0.00	0.25	0.13	0.00	0.62
Phenylethanol	1109	0.02	0.00	0.00	2.92	0.00	0.00	0.00
(3*E*,5*E*)-2,6-Dimethyl-1,3,5,7-octatetraene	1120	0.00	0.00	6.89	0.72	3.53	0.00	0.00
Alloocimene	1129	0.00	0.00	4.03	0.13	15.00	0.00	0.00
Benzyl cyanide	1139	2078.02	0.00	0.00	210.30	0.00	0.00	518.81
(4*E*,6*Z*)-2,6-Dimethyl-2,4,6-octatriene	1140	0.00	0.00	0.00	0.02	8.36	0.00	0.00
Pentyl/isopentyl 3-methylbutyrate & other unknown	1145	0.00	0.00	0.00	0.54	0.00	0.00	0.00
Benzyl acetate	1165	0.00	0.00	0.00	0.37	0.00	0.32	0.00
2-sec-Butyl-3-methoxypyrazine	1174	18.39	0.00	0.00	0.00	6.29	0.00	235.01
Unknown	1174	0.00	0.00	1.69	10.38	0.11	0.00	0.00
Naphthalene	1175	0.93	3.39	9.75	4.10	5.46	1.35	0.00
2-Methoxy-3-isobutylpyrazine	1181	32.99	0.00	0.00	0.00	4.09	0.00	0.00
Unknown	1184	2.40	0.00	0.84	0.00	3.26	6.87	0.00
Unknown	1185	0.75	0.00	10.05	0.00	5.74	0.00	0.00
Unknown ester	1188	0.00	0.00	0.00	4.65	0.09	1.29	0.00
Methyl salicylate	1189	1.01	1.74	5.90	1.52	0.33	3.60	0.09
Unknown	1198	0.00	0.00	71.27	0.34	2.13	0.00	0.00
β-Cyclocitral	1217	0.00	0.00	0.00	8.04	0.18	0.00	0.00
(*Z*)-3-Hexenyl isobutyrate	1237	459.58	8.88	0.00	84.10	0.18	219.32	0.00
Hexyl-3-methyl butyrate	1241	1501.33	0.00	0.00	138.19	0.24	0.24	0.00
Phenylacetaldhyde oxime	1261	2.81	0.00	0.00	1.74	0.00	0.00	0.23
Unknown	1265	0.00	40.65	0.00	0.05	0.00	11.32	0.00
Unknown	1271	4.29	0.00	0.00	1.07	0.00	0.00	0.00
3-Undecanone	1275	0.00	0.00	0.00	5212.91	1.08	0.00	0.00

1-Tridecene	1286	0.00	0.00	0.00	38.85	0.00	0.00	0.00
Dihydroedulan II	1290	58.67	37.69	27.07	48.53	62.82	70.79	13.38
Unknown aromatic	1299	15.95	0.00	0.00	590.01	0.06	0.00	0.00
Unknown	1312	2.81	0.00	0.00	62.16	0.04	6.19	0.00
Hexyl 3-methyl-2-butenoate	1321	8.64	0.00	0.00	7.17	0.00	0.00	0.00
Unknown	1324	0.00	0.00	0.00	0.68	0.00	2.11	0.00
Unknown	1337	0.00	0.00	0.00	11.25	0.00	0.00	0.00
Unknown	1358	10.20	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	1361	0.00	60.32	0.00	0.10	0.00	25.76	0.00
Unknown	1372	0.00	5.90	0.00	0.02	0.00	4.08	0.00
α-Copaene	1373	2.00	0.64	0.00	0.92	1.03	0.00	0.94
(*Z*)-3-Hexenyl hexanoate	1381	129.18	24.31	0.00	0.00	0.10	0.38	0.00
Unknown alkene	1382	0.00	0.00	0.00	1.07	0.01	0.00	0.00
(*Z*)-3-Hexyl hexanoate (branched?)	1386	542.38	884.69	0.00	0.00	0.00	197.88	0.00
Hexyl hexanoate	1386	804.85	1.39	0.00	0.05	0.01	0.00	0.00
Unknown	1390	8.56	6.42	0.00	2.12	4.85	0.00	37.05
Hexyl hexenoate	1390	1644.70	6.00	0.00	0.00	0.00	0.00	0.00
Unknown ester	1392	0.00	0.00	0.00	18.51	0.00	0.00	0.00
Unknown	1393	0.00	112.69	0.00	1.66	0.42	41.68	0.00
Unknown	1396	1419.45	736.73	418.50	236.94	452.03	0.00	1674.30
Unknown	1402	0.00	17.49	0.00	0.16	0.00	8.17	0.00
Unknown	1413	0.00	0.00	0.00	0.00	0.00	6.98	0.00
Unknown	1415	0.00	27.80	0.00	1.13	0.00	16.01	0.00
β-Caryophyllene	1417	4.87	27.33	0.00	5.34	3.81	1.34	0.00
Unknown	1423	13.27	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	1424	0.00	0.00	0.00	6.39	0.00	0.00	0.00
β-Copaene	1427	0.29	0.00	0.00	1.42	0.35	0.00	0.00
Unknown terpene	1427	0.00	0.00	0.00	0.79	0.00	1.07	0.00
Unknown aromatic	1433	27.96	24.21	0.00	9.72	18.46	0.00	97.33
Hexyl (*E*)-2-hexenoate	1436	40.40	0.00	0.00	0.03	0.00	0.00	0.00
7,8-Dihydro-β-ionone	1436	0.00	867.12	0.00	110.16	24.31	729.23	0.00
Unknown monoterpene	1438	151.66	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	1442	0.31	0.00	0.00	2.95	0.00	0.00	0.00
Unknown	1444	0.00	27.03	0.00	2.23	0.31	8.16	0.00
Unknown	1445	36.34	0.00	0.00	0.02	0.00	0.00	0.00
6,10-Dimethyl-5,9-undecadien-2-one	1449	0.00	0.00	0.00	5.79	0.00	0.00	0.00
Alloramadendrene	1458	0.00	0.00	0.00	1.53	0.32	0.00	0.00
11-Dodecanolide	1466	0.90	0.00	0.00	0.00	15.09	0.00	0.00
Unknown	1468	0.00	7.53	0.00	0.00	0.00	0.13	0.00
Unknown	1469	23.42	0.00	0.44	0.00	0.00	0.00	0.00
Unknown	1470	34.37	0.00	0.00	0.43	0.29	0.00	0.00
Unknown	1475	3.50	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	1478	7.88	0.00	0.00	0.00	0.00	0.00	0.00

Unknown hexenyl ester	1480	0.00	279.44	0.00	0.00	0.00	11.27	0.00
Germacrene	1486	4.01	1.79	0.20	20.33	6.40	0.87	0.00
Unknown	1488	0.16	0.00	0.00	1.58	0.00	0.00	0.00
Valencene	1492	2.58	0.00	0.00	2.06	0.00	0.00	0.00
5-Decanolide	1493	59.26	0.00	0.00	0.00	0.00	0.00	0.00
Pentadecane	1500	0.00	0.00	2.22	1.77	0.67	0.00	0.00
(E,E)-α-Farnesene	1511	19.52	0.00	36.63	47.08	5.37	0.00	0.00
12-Dodecanolide	1515	10.12	0.00	0.00	0.00	85.27	0.00	0.47
δ-Cadinene	1522	4.33	1.64	0.30	10.73	1.11	0.00	0.55
Unknown	1522	0.00	0.00	0.00	0.00	0.53	0.00	0.92
Homovanillyl alcohol	1523	0.00	0.00	0.14	0.06	1.33	0.00	2.86
Dihydroactinidiolide	1532	0.00	0.00	0.00	0.30	0.00	0.00	0.00
Unknown	1535	0.00	20.69	0.00	0.00	0.00	0.57	0.00
Calcorene	1540	0.00	0.00	0.26	0.06	0.06	0.00	0.00
Mellein	1540	0.00	0.61	0.00	170.43	0.02	0.00	0.00
Unknown	1543	0.43	0.14	0.09	0.01	0.02	0.00	1.34
(*E*)-Nerolidol	1563	0.86	24.21	0.51	5.89	0.28	0.00	0.00
Unknown	1565	0.00	0.00	0.00	45.28	0.00	0.00	0.00
Ester hexanoate	1566	14.57	0.00	0.00	0.00	0.06	0.00	0.00
(*Z*)-3-Hexenyl benzoate	1571	6.34	2.32	0.00	0.06	0.00	0.31	0.00
Unknown	1573	91.19	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	1576	3.31	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	1581	0.00	0.00	0.00	0.00	0.08	0.00	14.63
Unknown alkene	1581	0.75	0.00	0.00	0.73	0.00	0.00	0.00
Unknown	1583	31.61	0.00	0.00	0.00	0.00	0.00	0.00
Hexadecane	1600	0.00	0.00	5.43	2.94	1.30	0.00	0.00
Unknown	1607	0.00	0.00	0.00	0.00	4.17	0.00	0.00
1-*epi*-Cubenol	1608	0.00	0.00	0.00	0.57	0.01	0.00	0.00
Unknown	1621	0.00	0.00	0.00	17.27	0.00	0.00	0.00
Unknown	1625	2.54	5.39	0.00	0.00	0.00	0.00	0.00
Unknown	1630	24.71	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	1632	4.56	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	1636	1.01	0.00	4.74	0.00	3.85	0.00	0.00
Unknown	1639	0.00	0.00	2.69	1.01	0.61	0.00	0.00
Unknown	1643	3.59	0.00	0.00	0.00	0.00	0.00	0.00
δ-Cadinol	1644	0.36	0.37	0.00	23.23	0.14	0.00	0.00
Unknown	1646	0.00	0.00	0.00	0.00	16.03	0.00	0.00
α-Cadinol	1652	0.00	0.00	0.00	2.21	0.00	0.00	0.00
Unknown	1656	0.34	0.68	6.80	1.45	0.83	0.99	0.00
Unknown	1661	0.00	8.51	0.00	0.00	0.00	1.22	0.00
Heptadecene	1679	0.00	0.00	1.25	1.02	0.21	0.00	0.00
Pentadecanolide	1691	0.00	412.12	0.00	0.25	0.00	211.71	0.00
Heptadecane	1700	0.00	0.00	2.18	1.03	0.33	0.00	0.00
Unknown1	1704	0.00	0.00	0.00	0.00	16.03	0.00	0.00
Unknown2	1704	0.00	0.00	0.00	0.94	0.00	0.00	0.00
Unknown macrolide	1714	0.00	0.00	0.00	23.77	0.00	0.00	0.00

14-Tetradecanolide	1733	67.79	0.00	12.28	0.00	157.47	0.00	20.83
Unknown	1750	0.96	0.00	0.00	1048.23	0.01	0.00	0.00
Unknown hydrocarbon	1750	0.92	0.99	4.22	14.93	1.85	1.08	1.29
Unknown	1754	9.36	0.00	0.00	0.00	0.00	0.00	0.00
Tetradecanoic acid	1755	0.13	0.00	0.00	32.14	0.00	0.00	0.00
Ethyl benzoate	1762	0.33	0.00	0.38	0.49	0.48	0.00	0.37
Unknown ester	1765	39.52	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	1769	0.65	0.00	0.00	1.57	0.00	0.00	0.00
Unknown	1771	0.00	5.11	0.00	0.00	0.00	0.00	0.00
Unknown hexenyl or cyclo pentyl ester	1775	797.54	4.61	0.00	0.00	2.75	0.00	78.44
Unknown macrolide	1777	264.35	0.00	0.00	0.00	0.00	0.00	1.85
Unknown	1787	0.00	0.00	0.00	2.58	0.00	0.00	0.00
Octadecane	1800	0.00	0.00	2.34	0.75	0.45	0.00	0.00
Unknown	1802	0.00	0.00	0.00	1.16	0.00	0.00	0.00
Hexadecadienolide-15-olide	1807	0.00	0.00	0.00	28.73	0.00	0.00	0.00
Unknown	1817	0.25	6.33	0.00	0.00	0.00	0.00	0.50
Hexadecen-11-olide	1828	108.76	0.00	0.00	0.00	0.00	0.00	7.50
Unknown	1833	0.00	0.00	0.00	8.79	0.00	0.00	0.00
Hexadecenolide	1835	0.00	0.00	0.00	0.00	0.21	0.00	1.49
Neophytadiene	1836	0.00	315.81	0.00	0.13	0.00	21.27	0.00
Hexadecenolide	1845	100.43	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	1846	29.45	10.05	0.48	3.94	9.70	36.17	1.66
15-Hexadecanolide	1853	17.17	0.00	0.29	0.05	4.48	0.00	10.29
Hexanyl ester	1853	0.00	1489.91	0.00	0.00	0.25	150.00	0.00
Hexadecenolide	1854	24.29	0.00	0.00	0.00	0.00	0.00	0.20
Unknown	1856	21.35	0.00	0.00	0.00	1.40	0.00	5.16
Unknown	1860	6.31	0.00	0.00	0.00	2.13	0.00	13.14
Hexadecenolide	1861	441.07	0.00	7.61	0.00	3.61	0.00	63.47
Neophytadiene	1861	0.00	74.61	0.00	0.00	0.00	4.33	0.00
Benzyl salicylate	1870	1.36	6106.41	23.37	34.69	17.95	3322.00	0.00
Nonadecene	1877	1.08	0.00	1.38	0.84	0.13	0.00	0.00
Unknown macrolide	1878	0.00	54.84	0.00	0.00	0.00	1.08	0.00
Neophytadiene	1878	0.00	117.30	0.00	0.00	0.00	7.72	0.00
1-Hexadecanol	1884	0.00	0.00	0.00	11.29	0.21	0.00	0.00
Hexadecatrienolide	1884	11.33	0.00	0.00	0.00	0.00	0.00	1.06
9,11-Hexadecadien-11-olide	1893	115.51	0.00	0.00	0.00	0.00	0.00	2.63
Nonadecane	1900	51.07	0.00	25.48	0.30	10.28	0.00	28.87
Unknown	1901	1.47	0.00	4.28	1.67	0.01	0.00	2.92
Unknown sesquiterpene	1902	0.00	0.00	39.74	0.00	12.54	0.00	0.00
Unknown	1907	0.00	0.00	0.00	18.45	0.00	0.00	0.00
Unknown	1915	0.00	0.00	2.92	1.86	0.66	0.00	0.00
16-Hexadecanolide	1936	277.89	19.71	0.00	0.00	0.56	50.67	119.28
Hexadecen-16-olide	1937	8.02	1.06	0.00	0.00	0.50	0.00	2.24

Hexyl dodecanoate	1960	0.00	0.00	0.00	28.05	0.00	0.00	0.00
Hexadecanoic acid	1960	14.57	2.09	0.00	30.03	6.44	0.00	2.22
Unknown hydrocarbon	1962	0.14	0.00	0.86	0.23	0.41	0.00	0.16
Unknown	1968	0.17	0.00	4.11	1.04	4.14	0.00	2.49
Unknown	1970	0.00	0.00	0.00	110.51	0.09	0.00	0.00
Brassicalactone	1971	131.03	0.00	7.17	0.00	0.12	0.00	147.59
Unknown	1977	0.00	0.00	0.00	4.41	0.02	0.00	0.00
Octadecatrienolide	1987	21.38	0.00	0.00	0.00	0.00	0.00	15.07
Icosane	2000	25.98	0.00	25.91	0.02	36.62	0.00	41.62
Octadecen-11-olide	2008	183.49	0.00	0.00	0.00	0.00	0.00	57.35
Unknown	2017	0.00	0.00	0.00	101.74	0.00	0.00	0.00
Octadecadienolide	2019	119.80	18.57	4.53	0.00	90.17	0.00	738.77
Isopropyl palmitate	2020	34.95	0.00	0.00	0.06	0.00	0.00	62.12
Octadecadienolide	2033	45.14	1.93	0.00	0.00	12.42	1.94	1318.41
Unknown pentyl ester	2033	0.00	0.00	0.00	226.39	0.00	0.00	0.00
(*Z*)-9-Octadecen-11-olide	2038	8513.77	0.00	99.69	0.00	0.74	0.00	3257.65
2-Phenylethyl decanoate	2038	0.33	0.00	0.00	10.75	0.00	0.00	0.00
(*Z*)-9-Octadecen-13-olide	2044	5137.74	0.00	53.09	0.00	56.84	0.00	5916.34
Octadecanoic acid ester	2045	105.19	0.00	0.00	0.00	8.67	0.00	39.54
Unknown macrolide	2048	15.61	0.00	7.30	0.00	12.89	0.00	93.84
12-Octadecanolide	2055	119.92	0.00	0.00	0.00	0.00	0.00	12.06
Unknown	2055	0.00	0.00	0.00	5.68	0.00	0.00	0.00
(*Z*9,*E*11,*Z*15)-9,11,15-Octadecatrien-13-olide	2061	2354.90	0.00	0.00	0.00	0.41	0.00	614.24
Unknown	2062	107.29	0.00	0.00	0.00	0.00	0.00	0.00
(*E*)-Octadec-9-en-12-olide	2065	429.44	0.00	0.00	0.00	0.00	0.00	98.09
(*Z*9,*E*11)- 9,11-Octadecadien -13-olide	2065	22006.74	0.63	8.26	0.00	0.96	0.00	1063.22
Henicosene	2068	12.36	4.66	1313.29	6.98	65.72	0.00	7.26
Henicosene	2072	110.06	0.00	2430.30	53.95	149.32	9.61	100.88
Henicosene	2086	3.92	14.05	19.45	3.03	24.45	0.00	8.53
Unknown	2093	0.00	0.00	0.00	6.63	0.00	0.00	0.00
Octadecadienolide	2095	786.47	0.00	0.00	0.00	0.00	0.00	7.97
Henicosane	2100	1925.11	0.00	2265.48	19.39	3803.86	0.00	4272.72
Unknown terpene ester	2120	0.00	0.00	0.00	52.68	0.00	0.00	0.00
9-Octadecen-18-olide	2123	1114.19	10.97	48.20	0.06	25.86	18.23	1703.38
18-Octadecanolide	2135	532.01	3.25	0.00	52.57	12.35	0.00	217.04
Unknown terpene ester	2139	0.00	0.00	0.00	714.11	0.00	0.00	0.00
Unknown	2144	24.10	0.00	0.00	0.00	0.04	0.00	32.41
Unknown terpene ester	2150	0.00	0.00	0.00	2490.83	0.47	0.00	0.00
Unknown amide	2157	0.05	0.00	0.00	1.94	0.28	0.00	1.03
Ethyl oleate	2159	6.32	0.00	103.74	0.00	74.15	0.00	656.14
Unknown	2160	0.16	0.00	0.00	18.63	0.70	0.00	0.00
Octadecanoic acid	2161	5.98	0.00	0.00	1.57	1.78	0.00	2.31
Unknown	2166	0.16	0.00	0.00	246.29	0.29	0.00	0.00
Unknown	2168	0.00	0.00	0.00	56.48	0.05	0.00	27.60

Isopentyl octadecadienoate	2177	24.35	0.00	55.57	0.00	6.99	0.00	44.03
Octadecadienolide	2178	416.33	0.00	0.00	0.00	0.71	0.00	416.42
Butyl hexadecanoate	2186	0.00	0.00	0.00	1.52	48.32	0.00	395.49
Isopropyl oleate	2188	6455.01	0.00	536.83	0.00	247.13	0.00	9353.76
Docosane	2200	0.47	0.00	15.50	1.89	42.63	0.00	13.07
Unknown diterpene	2205	0.00	0.00	0.00	326.23	0.00	0.00	0.00
Unknown macrolide	2210	46.31	0.00	0.00	0.00	0.00	0.00	30.66
Icosen-13-olide	2218	715.91	0.00	1.38	0.00	0.00	0.00	76.47
Icosanal	2219	43.01	1.31	12.38	12.68	5.65	0.00	14.32
Isopropyl octadecanoate	2224	242.19	0.00	2.52	0.00	0.80	0.00	187.79
Unknown	2229	0.00	0.00	0.00	2.16	0.00	0.00	0.00
Unknown	2238	105.25	0.00	0.00	0.00	2.50	0.00	178.84
Isopropyl octadecadienolate	2241	2165.26	0.00	9.06	0.00	0.00	0.00	1339.28
2-Phenylethyl dodecanoate	2243	0.00	0.00	0.00	335.57	0.00	0.00	0.00
Unknown	2245	0.00	0.00	0.00	0.00	5.88	0.00	156.00
Icosadien-15-olide	2250	312.24	0.00	0.00	0.00	0.00	0.00	5.33
Isoprenyl palmitate	2254	0.55	0.00	0.00	5.47	0.63	0.00	0.00
Icosenol	2255	0.00	0.00	732.36	0.11	4.68	0.00	0.00
Unknown	2258	0.00	0.00	0.00	5.14	0.00	0.00	0.00
13-Icosanolide	2259	67.71	0.00	0.00	0.00	0.00	0.00	100.04
Isopropyl 9,12-octadecadienoate	2263	673.31	0.00	0.00	0.00	0.00	0.00	101.38
Unknown	2263	0.00	0.00	0.00	0.00	0.00	0.00	747.20
Tricosene	2270	383.95	2.22	898.75	21.47	152.97	7.71	64.87
Tricosene	2275	387.50	30.92	3461.39	223.41	268.81	36.72	247.60
Unknown	2279	0.00	0.00	0.00	106.05	0.00	0.00	0.00
Unknown	2280	0.00	0.00	0.00	3.03	0.00	0.00	0.00
Isopentyl ester	2285	0.00	0.00	34.49	0.00	5.54	0.00	1.71
Isobutyl oleate	2287	67.84	0.00	843.45	0.00	231.52	0.00	147.86
Unknown	2290	3.67	0.66	4.14	0.07	0.99	29.37	1.56
Unknown	2297	0.00	0.00	0.00	26.19	0.00	0.00	0.00
Tricosane	2300	110.63	4.50	125.94	56.30	355.85	0.41	229.22
Unknown	2305	47.55	0.00	43.17	0.00	77.52	0.00	453.19
Unknown terpene ester	2310	0.00	0.00	0.00	13314.02	6.20	0.00	0.91
Unknown	2320	66.77	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	2320	0.00	0.00	0.00	29.16	0.00	0.00	0.00
Unknown	2324	0.00	0.00	0.00	278.45	0.00	0.00	0.00
Unknown	2325	60.06	0.00	1.16	0.00	0.49	0.00	36.57
Unknown	2329	0.00	0.00	0.00	37.54	0.00	0.00	0.00
20-Icosanolide	2330	264.60	0.00	0.00	0.00	3.32	0.00	333.62
Unknown	2336	0.00	0.00	0.00	13.06	0.00	0.00	0.00
Unknown amide	2347	28.54	0.31	53.14	110.07	28.27	8.05	87.41
Docosadienolide	2347	33.41	0.00	0.00	0.00	0.14	0.00	25.95
Unknown	2349	0.00	0.00	0.00	7.66	0.00	0.00	0.00

Hexyl hexadecenoate	2351	36.26	0.00	0.00	0.47	0.96	0.00	0.00
Hexyl hexadecenoate	2353	66.90	5.49	0.00	14.14	8.16	0.00	0.00
Butyl octadecadienoate	2355	14.13	0.00	6.38	0.00	51.78	0.00	130.87
Unknown	2360	18.01	0.00	0.00	0.00	0.00	0.00	23.06
Butyl oleate	2361	0.00	0.00	74.72	18.11	1176.37	0.00	14387.66
Unknown sesterterpene hydrocarbon	2370	0.00	0.00	0.00	463.94	0.00	0.00	0.00
Unknown	2373	0.00	0.00	0.00	1.12	0.00	0.00	0.00
(*Z*)-3-Hexenyl hexadecanoate	2379	253.15	39.69	1.10	163.00	119.56	1.55	50.99
Butyl octadecanoate	2386	0.00	0.00	0.00	0.13	29.30	0.00	233.43
Unknown	2387	0.00	0.00	0.00	82.89	0.00	0.00	0.00
Unknown macrolide	2392	115.24	0.00	1.91	0.00	0.00	0.00	3.73
Unknown	2396	0.00	0.00	0.00	0.00	0.67	0.00	2287.35
Unknown	2398	0.00	0.00	0.00	27.11	0.00	0.00	0.00
Tetracosane	2400	0.25	0.00	14.73	12.99	10.04	0.00	0.81
Icosyl acetate	2408	0.00	0.00	0.00	53.34	0.00	0.00	0.00
2-Phenylethyl tetradecenoate	2415	0.00	0.00	0.00	7.07	0.00	0.00	0.00
Unknown	2418	80.06	0.00	0.00	39.32	2.59	0.00	236.97
19-Methylicosyl acetate	2422	73.80	0.46	5.47	5.31	9.16	0.00	0.00
Unknown sesterterpene hydrocarbon	2427	0.00	0.00	0.00	195.86	0.00	0.00	0.00
Unknown	2432	0.00	0.00	0.00	3.35	0.00	0.00	0.00
Unknown terpene ester	2435	0.00	0.00	0.00	1458.48	0.00	0.00	0.00
Isoprenyl octadec-11-enoate	2436	10.90	0.00	5.22	36.11	12.59	0.00	16.40
Unknown	2451	0.00	0.00	0.00	871.50	0.00	0.00	0.00
Unknown	2453	56.78	0.00	0.00	0.00	0.00	0.00	25.51
Docosenolide	2455	27.89	0.00	0.00	0.00	0.00	0.00	61.64
Unknown	2455	9.18	0.00	0.00	0.16	5.35	0.00	11.35
Unknown	2458	0.00	0.00	0.00	15.53	0.00	0.00	0.00
(Z)-13-Docosen-1-ol	2460	74.44	4.12	1423.15	2.06	66.81	0.00	46.98
Unknown	2460	0.00	0.00	0.00	14.00	2.34	0.00	0.00
Unknown	2465	8.67	1.07	0.00	0.00	1.91	0.00	6.27
Unknown	2467	12.46	0.00	3.47	0.00	0.71	12.63	0.00
Unknown	2472	66.43	0.00	47.08	29.73	224.34	0.00	431.21
Icosenolide	2475	59.95	0.00	0.00	0.00	0.00	0.00	105.64
Unknown	2487	0.00	0.00	0.00	375.56	0.00	0.00	0.00
1-Docosanol	2490	1606.60	26.14	2019.66	333.96	864.97	26.49	817.09
Unknown	2490	0.00	0.00	0.00	15.96	0.00	0.00	0.00
Unknown terpene ester	2494	0.00	0.00	0.00	1352.10	0.00	0.00	0.00
Pentacosane	2500	32.26	67.62	89.86	127.99	128.42	9.12	115.21
Unknown	2503	0.00	0.00	0.00	4.98	0.00	0.00	0.00
Unknown aromatic ester	2511	0.00	0.00	0.00	128.44	0.00	0.00	0.00
Unknown	2523	0.00	0.00	0.00	237.75	0.00	0.00	0.00
11-Methylpentacosane	2532	0.00	0.92	281.90	0.00	42.18	0.00	10.23

Hexyl octadecadienoate	2544	0.00	0.00	0.00	0.00	0.00	0.00	110.64
Hexyl octadecanoate	2550	50.62	182.27	1.10	55.23	97.19	29.26	40.25
Docosen-22-olide	2551	294.37	0.00	0.00	0.00	4.41	0.00	162.16
Hexyl octadecenoate & (*Z*)-3-Hexenyl octadecenoate	2557	7351.86	894.88	283.47	719.73	975.62	233.40	4569.69
Hexenyl octadecatrienoate & (*Z*)-3-Hexenyl octadecatrienoate	2561	1040.52	101.96	0.00	35.29	226.38	7.71	438.91
Benzyl hexadecanoate	2569	64.43	0.00	0.00	9.76	0.00	0.00	46.40
(*Z*)-3-Hexenyl octadecanoate	2581	194.51	4.37	2.36	16.92	82.54	0.66	26.03
Unknown	2588	0.00	0.00	0.00	163.90	0.00	0.00	0.00
Hexyl octadecanoate	2590	1304.10	0.00	0.00	0.00	0.00	0.00	417.94
Unknown	2593	0.00	0.00	0.00	402.39	0.00	0.00	0.00
Hexacosane	2600	0.00	7.30	14.07	27.73	20.78	0.69	25.86
1,3-Docosanediol (cyclic dImethylsilyl derivative)	2604	626.77	5.10	132.85	144.33	148.81	1.31	135.78
Hexyl octadecanoate	2621	113.97	0.00	0.00	0.00	0.62	60.76	386.24
Unknown	2626	0.00	0.00	5.05	0.00	2.25	0.00	0.00
Unknown	2628	0.00	0.00	0.00	58.75	0.00	0.00	0.00
Unknown	2631	0.00	22.63	0.52	3.16	0.01	0.18	0.00
Unknown sesterterpene	2636	0.00	0.00	0.00	5844.44	2.26	0.00	0.00
Unknown	2636	1.81	0.00	0.00	19.29	0.00	0.00	0.00
Unknown	2655	0.00	0.00	24.45	9.13	2.33	0.00	50.19
Unknown	2656	0.00	0.00	0.00	1.64	12.89	0.00	0.00
Unknown	2657	4.70	0.00	0.00	500.69	0.00	0.00	0.00
Tetracosenol	2664	2831.00	60.76	1491.35	747.98	1024.14	41.48	1536.26
Unknown	2670	0.00	0.00	3.92	37.23	3.22	0.00	0.00
Heptacosene	2675	18.98	4.46	20.89	38.51	85.04	0.00	168.58
Unknown	2690	91.66	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	2694	1345.00	0.00	0.00	177.23	171.59	0.00	478.60
1-Tetracosanol	2694	631.83	11.12	1574.08	801.80	806.30	13.84	578.68
Unknown	2699	0.00	0.00	0.00	436.84	0.00	0.00	0.00
Heptacosane	2700	323.89	237.58	238.02	222.08	280.33	139.08	318.56
1,3-Tricosanediol (cyclic dImethylsilyl derivative)	2704	99.89	0.00	0.00	0.00	0.00	0.00	0.00
Unknown aromatic ester	2718	0.00	0.00	0.00	65.07	0.00	0.00	0.00
11-Methylheptacosane	2730	0.00	48.91	42.01	0.00	28.38	2.56	65.08
Unknown	2733	5.38	14.02	0.00	62.42	0.00	0.00	37.76
Unknown	2735	0.00	0.00	0.00	489.73	0.00	0.00	0.00
Unknown2	2735	60.84	0.00	0.00	0.00	0.00	0.00	0.00
Cholestadiene	2744	32.82	7.42	1.03	6.29	7.37	0.00	17.60
Unknown	2746	0.00	0.00	2.07	0.31	1.29	0.00	0.00
Tetracosenolide	2749	1191.60	0.00	21.22	0.00	7.38	0.00	80.97
Unknown	2753	32.55	59.64	20.13	2.30	19.50	44.22	31.60
Unknown terpene	2755	0.00	0.00	0.00	1585.67	0.19	0.00	0.00
13-Docosenamide	2770	171.59	3.47	297.50	442.28	190.84	88.54	408.92

Unknown	2783	163.02	0.00	32.86	3.66	42.87	0.00	54.20
1,3-Tetracosanediol (cyclic dImethylsilyl derivative)	2799	302.17	1.02	229.22	155.63	149.58	0.00	87.12
Octacosane	2800	0.00	0.49	14.21	41.69	16.16	0.00	41.03
Unknown triterpene	2823	5.39	0.00	303.06	6449.52	46.90	0.96	19.88
Hexacosanal	2829	1270.51	6.38	120.75	759.60	200.84	0.72	124.36
Unknown	2840	27.15	0.00	0.00	1083.40	0.00	0.00	14.53
Unknown	2855	10.50	0.00	52.75	6.24	2.43	0.00	32.95
Unknown	2864	23.12	0.00	0.00	99.22	0.00	0.00	1.25
Cholesta-3,5-diene	2871	24.12	7.15	0.52	8.31	11.48	0.00	11.90
Unknown	2891	52.16	0.00	84.04	172.51	30.87	0.00	30.63
Unknown triterpene	2891	0.00	0.00	0.00	3126.08	0.99	0.00	0.00
Unknown	2898	376.41	0.00	28.91	54.49	186.54	0.00	123.42

Table S7: *Pairwise PERMANOVA p-values of between species androconial comparisons using Bonferroni correction. All values were significant at 0.05 significance level.*

	H. cydno	*H. eleuchia*	*H. elevatus*	*H. erato*	*H. melpomene*	*H. sapho*
H. eleuchia	0.021					
H. elevatus	0.042	0.021				
H. erato	0.021	0.021	0.021			
H. melpomene	0.021	0.021	0.021	0.021		
H. sapho	0.021	0.021	0.021	0.021	0.021	
H. timareta	0.021	0.021	0.021	0.021	0.021	0.021

Table S8: *Pairwise PERMANOVA p-values of between species genital comparisons using Bonferroni correction. Significant results are highlighted in bold.*

	H. cydno	*H. eleuchia*	*H. elevatus*	*H. erato*	*H. melpomene*	*H. sapho*
H. eleuchia	**0.021**					
H. elevatus	**0.021**	**0.021**				
H. erato	**0.021**	**0.021**	**0.021**			
H. melpomene	**0.021**	**0.021**	0.273	**0.021**		
H. sapho	**0.021**	**0.042**	**0.021**	**0.021**	**0.021**	
H. timareta	**0.021**	**0.021**	**0.021**	**0.021**	**0.021**	**0.021**

Table S9: *Summary of androconial chemical bouquet analysis of all species using the ManyGLM approach including all significant explanatory variables. The ten compounds which contribute the most to the deviance explained by a variable are listed for each variable in descending order of contribution. Compounds highlighted with * were also identified by an indicator analysis.*

Parameter	Residual DF	DF	Deviance	p-val	Compounds
Species	245	6	10944	0.001	Geranylgeranylacetone*, syringaldehyde, methyloctadecanal (RI=2076), icosanal, octadecanal*, (*Z*)-11-icosenal, henicosane, methyloctadecanal (RI=2064), unknown RI=1396, 1-hexadecanol
Region	242	3	3717	0.001	Henicosane, tricosane, unknown ester RI=1188, unknown RI=2133, tetracosane, napthalene[1], unknown RI=1366, unknown RI=2277, pentacosane
(Region/Locality)	231	11	2826	0.001	Hexadecadien-15-olide, unknown RI=1915, unknown hydrocarbon RI=1962, 1-hexadecanol, napthalene[1], nonanal, (*Z*)-13-docosenal, methyl 4-hydroxy-3-methoxybenzoate, (*Z*)-β-ocimene, unknown RI=1184
Species*Region	222	9	896	0.005	Napthalene, methyl salicylate, henicosane, 1-octadecanol, mellein, dihydroactinidiolide, 1-hexadecanol, octadecanal, (*Z*)-13-docosen-1-ol, tricosane
Species*(Region/Locality)	220	21	741	0.001	1-Hexadecanol, pentacosane, 1-octadecanol, methyl salicylate, henicosane

[1]Naphthalene is a known flower volatile, but can also be introduced by contamination. Our blank samples never contained naphthalene, indicating the butterfly origin in our study.

Table S10: *Summary of genital chemical bouquet analysis of all species using the ManyGLM approach including all significant explanatory variables. The ten compounds which contribute the most to the deviance explained by a variable are listed for each variable. Compounds highlighted with * were also identified by an indicator analysis.*

Parameter	Residual DF	DF	Deviance	p-val	Compounds
Species	268	6	27587	0.001	Unknown terpene ester RI=2494*, unknown terpene ester RI=2139, henicosane, unknown pentyl ester RI=2033, unknown terpene ester 2435, unknown aromatic RI=1299, unknown terpene RI=2755, benzyl cyanide, unknown sesterterpene hydrocarbon RI=2370, (*E*)-β-ocimene*
Region	265	3	8965	0.001	Unknown RI=2840, 7,8-dihydro-β-ionone, benzyl cyanide, unknown aromatic ester RI=2511, 2-phenylethyl dodecanoate, unknown triterpene RI=2891, unknown aromatic ester RI=2718, hexadecane, unknown RI=1076, 3-undecanone
(Region/Locality)	257	11	6431	0.016	(*Z*)-β-Ocimene, unknown RI=1915, unknown hydrocarbon RI=1750, 18-octadecanolide, henicosene (2068), 19-methylicosyl acetate, napthalene, hexadecanoic acid, 3-undecanone, unknown terpene ester RI=2310

Table S11: *Androconial and genital compounds which are the best indicators of different geographic groups of* H. erato*. A is a measure of group specificity of the compounds, B is a measure of group coverage, and sqrtIV is the indicator value which considers both A and B and ranges from 0 (compound not present in any individuals of that species) to 1 (compound only present in that species, and present in all individuals).*

Wings			
Amazon	**A: specificity**	**B: coverage**	**sqrtIV**
Napthalene	0.787	0.870	0.827
East Andes			
1-Hexadecanol & mellein	0.802	0.957	0.876
West Andes			
Unknown RI=1704	0.950	0.833	0.890
Panama			
Benzylacetate	1	1	1
Unknown ester RI=1188	1	1	1
Genitals			
Amazon			
Napthalene	0.851	0.957	0.902
East Andes			
Unknown triterpene RI=2891	0.846	0.978	0.910
West			
Unknown RI=1833	1	1	1
Unknown RI=1970	1	1	1
2-Phenylethyl decanoate	1	1	1
Unknown terpene ester RI=2120	1	1	1
2-Phenylethyl dodecanoate	1	1	1
Unknown RI=2258	1	1	1
2-Phenylethyl tetradecenoate	1	1	1
Unknown aromatic ester RI=2511	1	1	1
Unknown aromatic ester RI=2718	1	1	1
Unknown RI=2734	1	1	1
Panama			
Benzyl acetate	1	1	1
Unknown ester RI=1188	1	1	1
Pentyl/isopentyl 3-methylbutyrate	1	1	1

Table S12: *Androconial and genital compounds which are the best indicators of different geographic groups of* H. melpomene. *A is a measure of group specificity of the compounds, B is a measure of group coverage, and sqrtIV is the indicator value which considers both A and B and ranges from 0 (compound not present in any individuals of that species) to 1 (compound only present in that species, and present in all individuals).*

Wings			
Amazon	**A: specificity**	**B: coverage**	**sqrtIV**
Alkene or alcohol (RI=2127) & henicosane	0.871	0.966	0.917
East Andes			
(*Z*)-13-Docosenal & henicosane	0.962	0.895	0.928
West Andes			
Unknown RI=1766	1	0.727	0.852
Panama			
Nonanoic acid	0.816	1	0.903
Genitals			
Amazon			
14-Tetradecanolide	0.915	0.975	0.945
East Andes			
7,8-Dihydro-β-ionone	1	0.881	0.939
West			
Hexyl octadecenoate (RI=2621)	0.826	1	0.909
Panama			
2-*sec*-Butyl-3-methoxypyrazine	0.495	1	0.704

Table S13: *Summary of androconial chemical bouquet analysis of* H. erato *using the ManyGLM approach including all significant explanatory variables. The ten compounds which contribute the most to the deviance explained by a variable are listed for each variable. Compounds highlighted with * were also identified by an indicator analysis.*

Parameter	Residual DF	DF	Deviance	p-val	Compounds
Region	84	3	1275	0.001	Pentacosane, unknown ester RI=1188*, benzyl acetate*, unknown macrolide RI=1714, tetracosane, unknown RI=1366, pentyl/isopentyl 3-ethylbutyrate RI=1145, mellein*, hexadecadien-15-olide, heptadecene,
(Region/Locality)	78	9	1091	0.001	Hexadecadien-15-olide, napthalene*, unknown RI=1915, unknown hydrocarbon RI=1962, tetracosane, unknown RI=1184, unknown macrolide RI=1714, pentacosane, unknown RI=1444, unknown RI=1424

Table S14: *Summary of genital chemical bouquet analysis of* H. erato *using the ManyGLM approach including all significant explanatory variables. The ten compounds which contribute the most to the deviance explained by a variable are listed for each variable. Compounds highlighted with * were also identified by an indicator analysis.*

Parameter	Residual DF	DF	Deviance	p-val	Compounds
Region	88	3	6142	0.001	3-Undecanone, unknown sesterterpene RI=2636*, unknown RI=2840, unknown tripterpene RI=2891*, unknown aromatic ester RI=2511*, 2-phenylethyl dodecanoate*, unknown RI=2451, unknown terpene ester RI=2435, unknown aromatic ester RI=2718*, benzyl cyanide,
(Region/Locality)	82	9	2721	0.001	Napthalene*, hexadecanoic acid, unknown diterpene RI=2205, unknown ester hexanoate RI=1565, 18-octadecanolide, unknown RI=2279, unknown macrolide RI=1714, unknown RI=1424, unknown amide RI=2157, icosanal

Table S15: *Summary of androconial chemical bouquet analysis of* H. melpomene *using the ManyGLM approach including all significant explanatory variables. The ten compounds which contribute the most to the deviance explained by a variable are listed for each variable.*

Parameter	Residual DF	DF	Deviance	p-val	Compounds
Region	83	3	1848	0.001	Henicosane*, tricosane, methyl 3,4-dimethoxybenzoate, homovanillyl alcohol, (*Z*)-13-docosen-1-ol*, 11-icosenol, icosenol, napthalene, unknown aromatic RI=1738, unknown alkene or alcohol RI=2127*
(Region/Locality)	75	11	1582	0.001	Unknown RI=2133, nonanal, 1-octadecanol, (*Z*)-13-docosenal, (*Z*)-9-octadecenal, Unknown RI=1915, (*Z*)-16-methyl-9-octadecenol, unknown RI=2112, unknown RI=1638, tricosene RI=2072

Table S16: *Summary of genital chemical bouquet analysis of* H. melpomene *using the ManyGLM approach including all significant explanatory variables. The ten compounds which contribute the most to the deviance explained by a variable are listed for each variable. Compounds highlighted with * were also identified by an indicator analysis.*

Parameter	Residual DF	DF	Deviance	p-val	Compounds
Region	100	3	2281	0.001	7,8-Dihydro-β-ionone*, 12-dodecanolide, 2-*sec*-butyl-3-methoxypyrazine, napthalene, unknown RI=1704, 2-methoxy-3-isobutylpyrazine*, (*Z*)-β-ocimene, unknown RI=1607, 11-dodecanolide, (*E*)-β-ocimene
(Region/Locality)	93	10	1777	0.001	12-Dodecanolide, unknown hydrocarbon RI=1750, unknown RI=1915, benzyl salicylate, hexenyl octadecatrienoate & (Z)-3-hexenyl octadecenoate, 11-methylpentacosane, unknown RI=2891, 14-tetradecanolide*, unknown sesquiterpene RI=1902, 11-dodecanolide

Chapter 5

A novel terpene synthase produces an anti-aphrodisiac pheromone in the butterfly *Heliconius melpomene*

Abstract

Terpenes, a group of structurally diverse compounds, are the biggest class of secondary metabolites. While the biosynthesis of terpenes by enzymes known as terpene synthases (TPSs) has been described in plants and microorganisms, few TPSs have been identified in insects, despite the presence of terpenes in multiple insect species. Indeed, in many insect species, it remains unclear whether terpenes are sequestered from plants or biosynthesised *de novo*. No homologs of plant TPSs have been found in insect genomes, though insect TPSs with an independent evolutionary origin have been found in Hemiptera and Coleoptera. In the butterfly *Heliconius melpomene*, the monoterpene (*E*)-β-ocimene acts as an anti-aphrodisiac pheromone, where it is transferred during mating from males to females to avoid re-mating by deterring males. To date only one insect monoterpene synthase has been described, in *Ips pini* (Coleoptera), and is a multifunctional TPS and isoprenyl diphosphate synthase (IDS). Here, we combine linkage mapping and expression studies to identify candidate genes involved in the biosynthesis of (*E*)-β-ocimene. We confirm that *H. melpomene* has two enzymes that exhibit TPS activity, and one of these, HMEL037106g1 is able to synthesise (*E*)-β-ocimene *in vitro*. Unlike the enzyme in *Ips pini,* these enzymes only exhibit residual IDS activity, suggesting they are more specialised TPSs, akin to those found in plants. Phylogenetic analysis shows that these enzymes are unrelated to previously described plant and insect TPSs. The distinct evolutionary origin of TPSs in Lepidoptera suggests that they have evolved multiple times in insects.

Significance statement

Terpenes are a diverse class of natural compounds, used by both plants and animals for a variety of functions, including chemical communication. In insects it is often

unclear whether they are synthesised *de novo* or sequestered from plants. Some plants and insects have converged to use the same compounds. For instance, (*E*)-β-ocimene is a common component of floral scent and is also used by the butterfly *Heliconius melpomene* as an anti-aphrodisiac pheromone. We describe two novel terpene synthases, one of which synthesises (*E*)-β-ocimene in *H. melpomene,* unrelated not only to plant enzymes but also other recently identified insect terpene synthases. This provides the first evidence that the ability to synthesise terpenes has arisen multiple times independently within the insects.

Introduction

Plants and insects sometimes use the same compounds for communication (Schiestl 2010; Beran et al. 2019). This may be adaptive if these chemicals exploit pre-existing sensory traits in the intended receiver. For example, sexually deceptive orchids mimic the scent of females of the pollinator species to attract males for pollination (Ayasse et al. 2011). Similarly, insects may use plant-like volatiles to exploit the sensory systems of other insects whose sensory systems have evolved for plant-finding (Baker 1989; Schiestl 2010; Conner and Iyengar 2016). Phenotypic convergences such as these may involve different molecular mechanisms, including independent evolution at different loci, or may be due to the exchange of genes through horizontal gene transfer (Stern 2013), and the concept has been studied across a range of organisms and phenotypes. However, we know little about the genetic basis of convergence in chemical signals.

One example of chemical convergence between plants and insects is the use of β-ocimene, a very common plant volatile, important in pollinator attraction due to its abundance and ubiquity in floral scents (Farré-Armengol et al. 2017). This compound is also found in the genitals of male *Heliconius* butterflies (Schulz et al. 2008; Estrada et al. 2011; Merrill et al. 2015). In *Heliconius melpomene*, (*E*)-β-ocimene acts as an anti-aphrodisiac pheromone, transferred from males to females during mating to repel further courtship from subsequent males (Schulz et al. 2008). β-Ocimene is also found in large amounts in the flowers on which adult *H. melpomene* feed, and elicits a strong antennal response in both males and females (Andersson et al. 2002; Andersson and Dobson

2003). This compound, therefore, appears to be carrying out two context-dependent functions, attraction to plants and repulsion from mated females.

Although β-ocimene synthases have been described in plants, none have been found in animals (Farré-Armengol et al. 2017). It has previously been shown that *H. melpomene* is able to synthesise (*E*)-β-ocimene *de novo* (Schulz et al. 2008). β-Ocimene is a monoterpene, a member of the largest and most structurally diverse class of natural products, the terpenes (Gershenzon and Dudareva 2007). Terpenes are formed from two precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are themselves produced by either the universal mevalonate pathway or the methylerythritol phosphate pathway, the latter of which is absent in animals (Eisenreich et al. 2004; Chen et al. 2016). Varying numbers of IPP units are then added to DMAPP to form isoprenyl diphosphates of different chain lengths by isoprenyl diphosphate synthases (IDSs) (Kellogg and Poulter 1997; Wang and Ohnuma 2000) (Fig. 1). These isoprenyl diphosphates are the precursors for the production of terpenes by terpene synthases (TPSs), with the length of the isoprenyl diphosphate determining the type of terpene that is made (Bohlmann et al. 1998; Christianson 2006). For example, DMAPP and one unit of IPP are converted by the IDS geranyl pyrophosphate synthase (GPPS) to form geranyl pyrophosphate (GPP), which can be converted by TPSs to monoterpenes, such as β-ocimene (Fig. 1). DMAPP and two units of IPP are converted by the IDS farnesyl diphosphate synthase (FPPS) into farnesyl diphosphate (FPP), which can be converted by TPSs to sesquiterpenes (Degenhardt et al. 2009) (Fig. 1). Finally, DMAPP and three units of IPP are converted by the IDS geranylgeranyl diphosphate synthase (GGPPS) into geranylgeranyl diphosphate (GGPP), which can be converted by TPSs to diterpenes (Fig. 1).

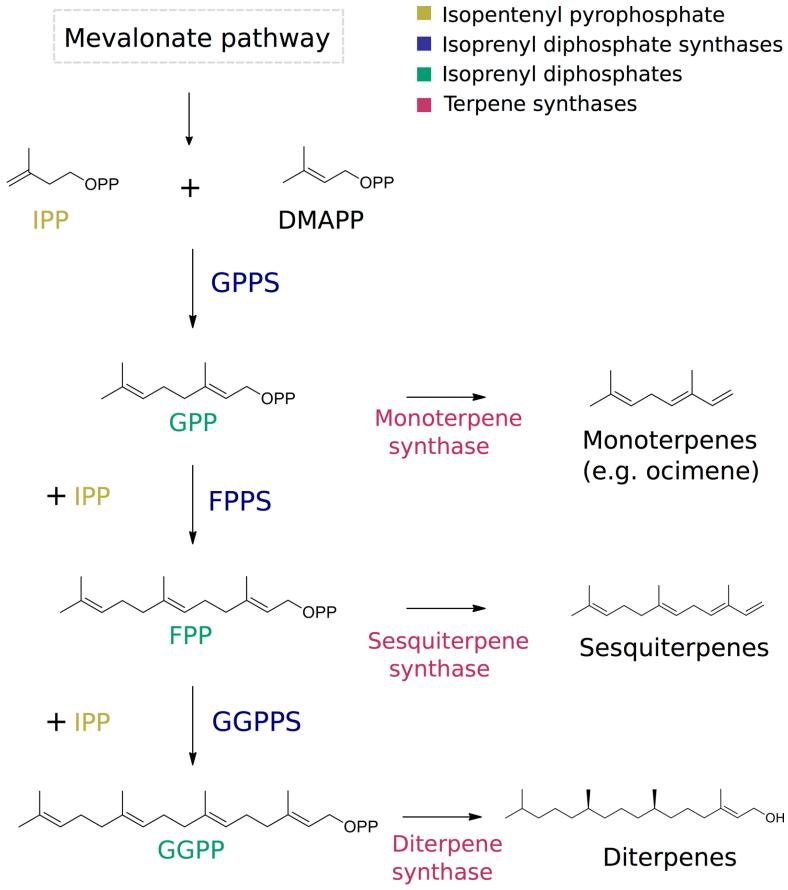


Figure 1: *Pathway of terpene biosynthesis. Isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are first formed from the mevalonate pathway. IPP and DMAPP are the substrates for isoprenyl disphosphate synthases which produce isoprenyl diphosphates of varying lengths, depending on the number of IPP units added. Isoprenyl diphosphates are themselves the substrates used by terpene synthases to make terpenes of various sizes, for example, monoterpene synthases produce monoterpenes, such as ocimene, from geranyl pyrophosphate (GPP). For illustration, (*E,E*)-α-farnesene is used as a representative sesquiterpene, and phytol as a diterpene.*

Both the mevalonate pathway, which forms IPP and DMAPP, and IDSs are ubiquitous in nature. In insects, the production of juvenile hormones is reliant on this pathway via FPP (Bellés et al. 2005). In contrast, TPSs are more limited in their distribution. Until recently they had only been described in plants and fungi in the eukaryotic domain, suggesting that insects sequestered terpenes from their diet and were unable to synthesise these compounds *de novo* (Chen et al. 2016). In the last decade, insect TPS genes, which are not homologous to plant TPSs, have been discovered in Hemiptera and Coleoptera, and were shown to be involved in the production of aggregation and sex pheromones (Gilg et al. 2005, 2009; Beran et al. 2016, 2019; Lancaster et al. 2018, 2019). The enzymes found in Hemiptera are involved in the production of pheromone precursor sesquiterpenes from FPP, although the enzymes catalysing the terminal pheromone biosynthesis steps are unknown (Lancaster et al. 2018, 2019). Sesquiterpene synthases have also been described in *Phyllotreta striolata* (Coleoptera) (Beran et al. 2016). The only monoterpene synthase described to date is that of *Ips pini* (Coleoptera), which produces a pheromone precursor from GPP (Gilg et al. 2005, 2009). These TPS genes have evolved from IDS-like genes, most closely related to FPPSs (Beran et al. 2016, 2019). The TPS of *Ips pini* also retains IDS function, acting as both a GPPS and TPS *in vitro*. It is unclear whether the evolution of TPS activity occurred only once in insects, as the most recent phylogenetic evidence suggests, or has occurred independently in different lineages (Beran et al. 2019; Lancaster et al. 2019).

Here, we identify the genes involved in the biosynthesis of (*E*)-β-ocimene in the butterfly *Heliconius melpomene* and analyse the evolution of terpene synthesis in *Heliconius* and other insects. To determine candidate TPS genes, we identified pathway orthologs in *H. melpomene* and carried out a genetic mapping study between *H. melpomene* and *H. cydno*, a closely-related species that does not produce (*E*)-β-ocimene. We identified a genomic region associated with the production of (*E*)-β-ocimene and searched for candidates within this region. We then identified genes with upregulated expression in the genitals of male *H. melpomene*, where (*E*)-β-ocimene is produced. We confirmed the TPS function of our candidate genes by expression in *E. coli* followed by enzymatic assays.

Results

Expansion of IDSs in genome of H. melpomene

We identified candidates potentially involved in terpene synthesis by searching in the genome of *H. melpomene* for enzymes in the mevalonate pathway and isoprenyl diphosphate synthases (IDSs) using well-annotated *Drosophila melanogaster* orthologs (Table S1)(Lai et al. 1998; Bellés et al. 2005; Noriega et al. 2006). We identified reciprocal best blast hits for all enzymes, except for acetoacetyl-CoA thiolase (Fig. 2). There was a clear one to one relationship for all enzymes, except for the IDSs which showed evidence for gene duplication. Of these, *Heliconius* contains two putatuive farnesyl diphosphate synthases (FPPSs), four putative copies of decaprenyl pyrophosphate synthase (DPPS) subunit two, and six putative geranylgeranyl pyrophosphate synthases (GGPPSs) (Fig. 2).

The biggest expansion found was that of the GGPPSs, which are IDSs that catalyse the addition of IPP to FPP to form GGPP. One of these, *HMEL015484g1*, shows 83% amino acid sequence similarity to the GGPPS of the moth *Choristoneura fumiferana*, which has previously been characterised *in vitro* to catalyse the production of GGPP from FPP and IPP (Barbar et al. 2013). *HMEL015484g1* is also the best reciprocal blast hit with the GGPPS of *D. melanogaster* (Fig. 2). The other five annotated GGPPSs show less than 50% similarity to the moth GGPPS, such that their function is less clear.

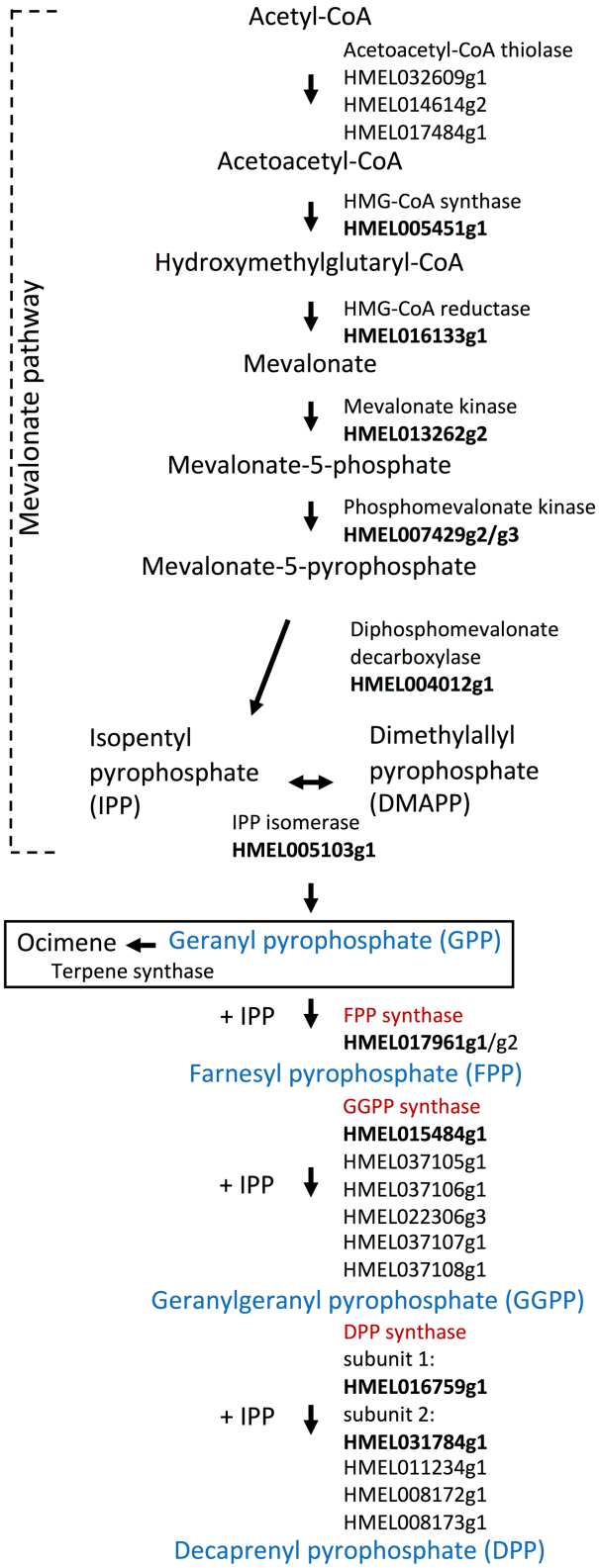


Figure 2: *Proposed biosynthetic pathway in* H. melpomene*. Reciprocal best blast hits are highlighted in bold. IDSs are in red and their products, IDs, in blue. The first two exons of* HMEL007429g2 *and the last exon of* HMEL007429g3 *are expressed as a single transcript (for transcript sequence see Table S1).*

QTL for (E)-β-ocimene production on chromosome 6

In order to determine which of the genes identified above could be important for (*E*)-β-ocimene production in *H. melpomene* we took advantage of the fact that a closely related species, *H. cydno*, does not produce (*E*)-β-ocimene (Fig. 3A). These two species can hybridise and, although the F1 females are sterile, F1 males can be used to generate backcross hybrids. We bred interspecific F1 hybrid males and backcrossed these with virgin females of both species to generate a set of backcross mapping families. The (*E*)-β-ocimene phenotype segregated in families backcrossed to *H. cydno* and so we focused on these families (Figure S1). Using quantitative trait locus (QTL) mapping with 114 individuals we detected a single significant peak on chromosome six associated with (*E*)-β-ocimene quantity (Fig. 3B). The QTL peak was at 36.4 cM, and the associated confidence interval spans 16.7-45.5 cM, corresponding to a 6.89Mb region containing hundreds of genes. The percentage of phenotypic variance explained by the peak marker is 16.4%, suggesting additional loci and/or environmental factor also contribute to the phenotype (Figure S2).

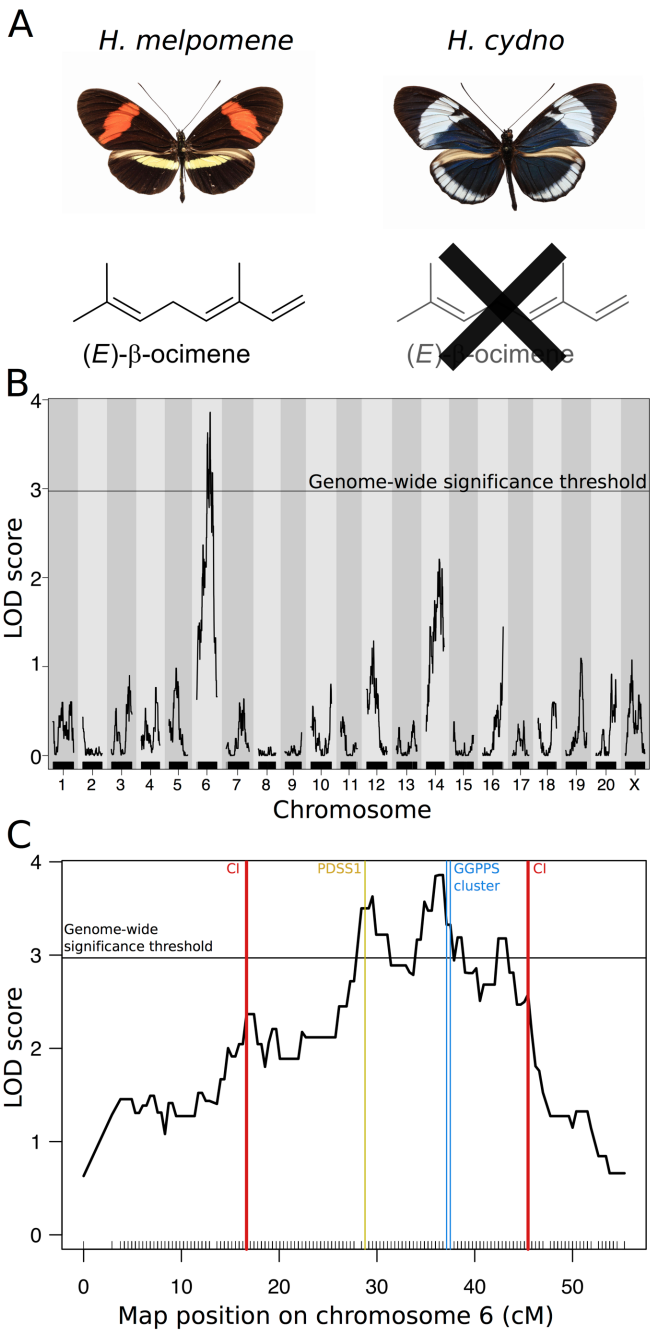


Figure 3: *QTL for (*E*)-β-ocimene production. A) The two species used in the crosses,* H. melpomene *which produces (*E*)-β-ocimene, and* H. cydno *which does not. B) Genome-wide scan for QTL underlying (*E*)-β-ocimene production. C) QTL on chromosome 6 for (*E*)-β-ocimene production. Confidence intervals (CI) as well as the positions of candidate genes (subunit 1 of decaprenyl diphosphate synthase (PDSS1) and the GGPPS cluster) in the region are marked. Black lines above x-axis represent the position of genetic markers and horizontal line shows genome wide significance threshold (alpha=0.05, LOD=2.97).*

Patterns of gene expression identify HMEL037106g1 *and* HMEL037108g1 *as candidates*

To identify candidate genes for (*E*)-β-ocimene production we searched within the confidence interval of the QTL peak. We found that subunit 1 of DPPS, as well as all six GGPPSs were found in this region (Fig. 3C). We then compared the expression levels of the seven genes found within the QTL using published RNA sequencing (RNA-seq) data (Walters et al. 2015). We first analysed data from *H. melpomene* male and female abdomens and heads, mapped to the *H. melpomene* reference genome. Since (*E*)-β-ocimene is found in male abdomens in *H. melpomene*, we hypothesised that its synthase would be highly expressed in this sex and tissue. Only one gene showed male abdomen-biased expression: *HMEL037106g1* (Fig. 4A, Table S3, sex*tissue, t=-4.35, p<0.01). All other genes did not show a significant bias in this direction (Fig.4A, Table S3).

We next compared gene expression between *H. cydno* and *H. melpomene* abdomens. If HMEL037106g1 is synthesising (*E*)-β-ocimene, we expect its expression to be higher in *H. melpomene* male abdomens than in *H. cydno*, given that *H. cydno* does not produce the compound. We generated a reference-guided assembly of *H. cydno* by aligning an existing *H. cydno* Illumina trio assembly (Malinsky et al. 2016), to the *H. melpomene* reference, followed by automated gene annotation (see Supplementary Information). We then manually identified *H. cydno* orthologs for our seven candidate genes and checked for differential expression between species and sexes. *HMEL037106g1* and *HMEL037108g1* were the only genes showing greater male-biased expression in *H. melpomene* abdomens than in *H. cydno* abdomens (Fig. 4B, Table S4, *HMEL037106g1*, species*sex, t=3.15, p=<0.05; *HMEL037108g1*, species*sex, t=3.44, p<0.05). No other genes showed a significant bias in this direction (Fig. S3, Table S4). In summary, *HMEL037106g1* and to a lesser extent *HMEL037108g1* are primary candidate genes from within the QTL region.

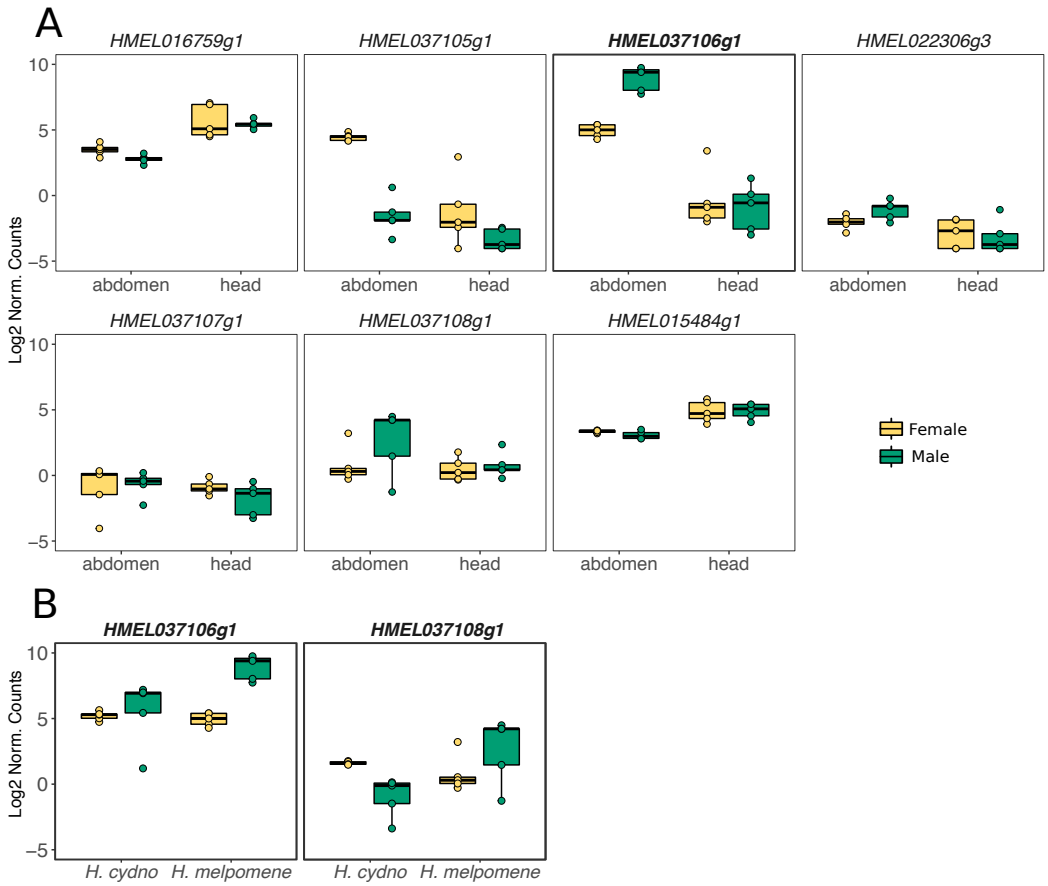


Figure 4: *Gene expression analysis of candidate genes. A) Expression of genes in* H. melpomene *heads and abdomens of males and females. HMEL037106g1 (highlighted) shows male abdomen-biased expression. B) Expression of genes in* H. melpomene *and H. cydno abdomens of males and females (expression of other genes in Fig. S3). Both HMEL037106g1 and HMEL037108g1 (highlighted) show greater male-biased expression in* H. melpomene *than* H. cydno*. Full model statistics in Table S3 and S4. N=5 for each boxplot. Gene expression is given in log2 of normalised counts (using the TMM (trimmed mean of M values) transformation).*

Functional experiments demonstrate the TPS activity of HMEL037106g1 and HMEL037108g1

We cloned HMEL037106g1 and HMEL037108g1 into plasmids and were able to generate heterologous expression of both proteins in *Escherichia coli*. We then conducted enzymatic assays with the expressed proteins using precursors from different points in the pathway to characterise their enzymatic function (Fig. 1, 5).

Firstly, we carried out assays with DMAPP and IPP, the two building blocks at the beginning of the terpene synthesis pathway to test for both IDS and TPS activity, as was seen in *Ips pini* (Fig. 1). HMEL037106g1 produced trace amounts of (*E*)-β-ocimene, linalool, another monoterpene, and nerolidol, a sesquiterpene, in this assay. This presumably occurs via the production of GPP and FPP, therefore HMEL037106g1 exhibits residual GPS and FPPS activity, as well as monoterpene synthase and sesquiterpene synthase activity to convert the GPP and FPP to (*E*)-β-ocimene, linalool, and nerolidol (Fig. 5A, Table S6). HMEL037108g1 produced trace amounts of linalool (Fig 5C) and nerolidol from DMAPP and IPP. Again, this demonstrates residual GPS and FPPS activity to form the GPP and FPP, and then both monoterpene and sesquiterpene synthase activity to convert these to linalool and nerolidol (Fig. 5A, Table S7).

We then carried out assays with GPP and IPP, as well as GPP alone to test for monoterpene synthase activity (Fig. 1). HMEL037106g1 showed monoterpene synthase activity, producing (*E*)-β-ocimene when provided with either GPP and IPP, or GPP alone (Fig. 5A, Table S6). Small amounts of (*Z*)-β-ocimene were also produced in treatments where (*E*)-β-ocimene was produced in large quantities (Table S6). In contrast to HMEL037106g1, HMEL037108g1 only produced (*E*)-β-ocimene in very small amounts from GPP (Table S9). Instead, linalool was produced in large amounts from GPP, suggesting that this enzyme is also acting as a monoterpene synthase but is responsible for production of linalool rather than (*E*)-β-ocimene (Fig. 5A, Table S7). HMEL037106g1 also produced linalool, albeit in much smaller quantities (Fig. 5A, Table S6).

Finally, we carried out assays with FPP and IPP to test for sesquiterpene synthase activity (Fig. 1). Although HMEL037106g1 exhibited small amounts of sesquiterpene

synthase activity through the trace production of nerolidol from DMAPP and IPP (Table S6), when provided with FPP alone, sesquiterpene synthase activity was not demonstrated, suggesting it is not the primary enzyme function (Fig. 5A, Table S6). In contrast, HMEL037108g1 did exhibit sesquiterpene synthase activity, producing large amounts of nerolidol when FPP was provided as a precursor (Fig. 5C, Table S7).

Due to the linalool detected in treatments where (*E*)-β-ocimene was produced by HMEL037106g1, we tested whether linalool could be a metabolic intermediate between GPP and (*E*)-β-ocimene. However, HMEL037106g1 did not produce (*E*)-β-ocimene from linalool (Fig. S5, Table S8). The two stereoisomers of linalool, (*S*)-linalool and (*R*)-linalool, have different olfactory properties. We confirmed the stereochemistry of linalool produced by both enzymes and found that whilst HMEL037106g1 produced mainly (*S*)-linalool, HMEL037108g1 produced a racemic mixture (Fig. S6).

In summary, HMEL037106g1 is a monoterpene synthase, catalysing the conversion of GPP to (*E*)-β-ocimene (Fig. 5C, Table S9). HMEL037108g1 is a bifunctional monoterpene and sesquiterpene synthase catalysing the conversion of GPP to linalool as well as FPP to nerolidol (Fig. 5C, Table S9).

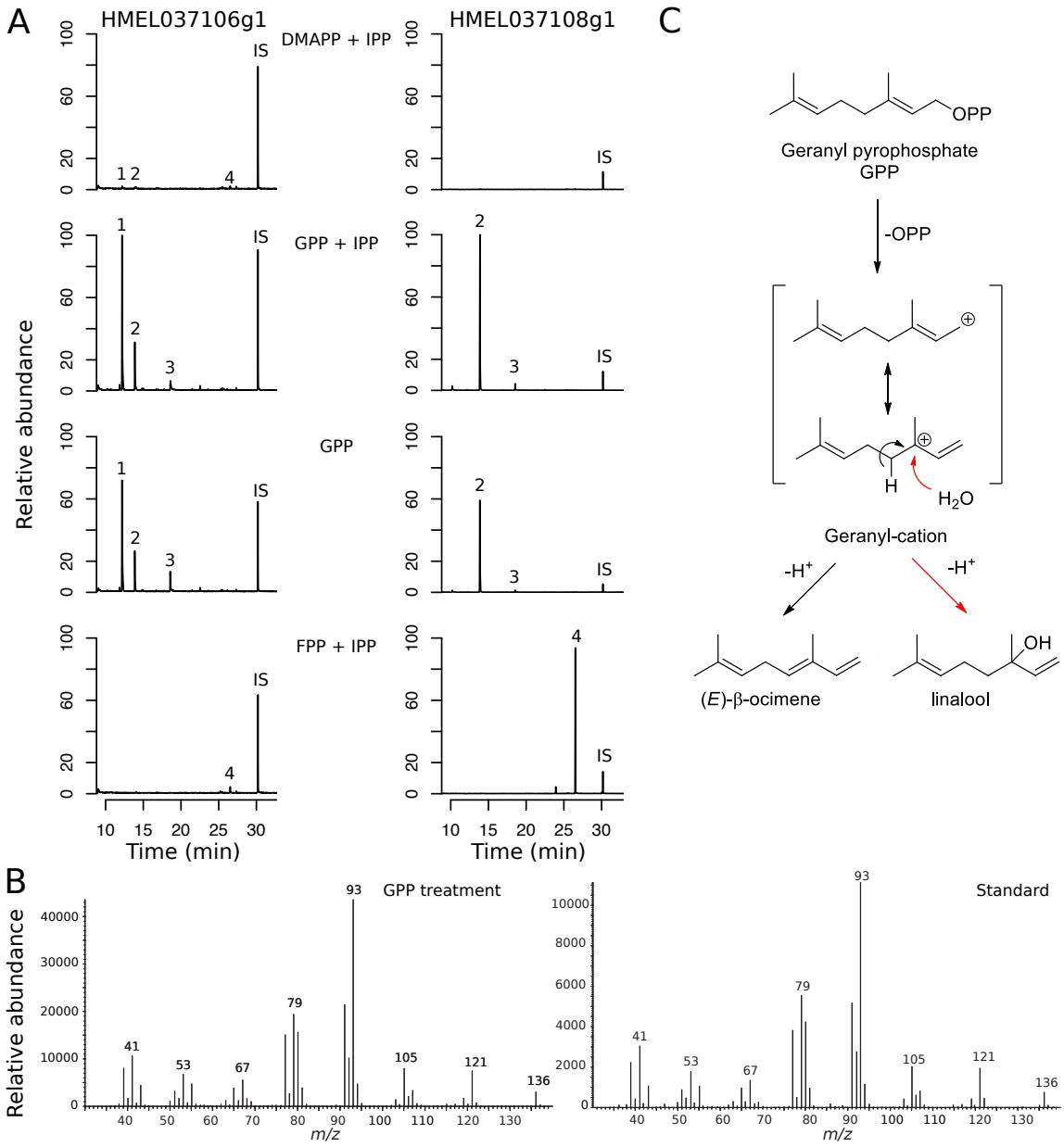


Figure 5: *Functional characterisation of TPS activity of HMEL037106g1 and HMEL037108g1 from* H. melpomene*. A) Total ion chromatograms of enzyme products in the presence of different precursor compounds. HMEL037106g1 produces high amounts of (*E*)-β-ocimene in the presence of GPP, with trace amounts found in the treatment with DMAPP + IPP, and none with FPP. HMEL037108g1 produces large amounts of linalool with GPP, and nerolidol with FPP. 1, (*E*)-β-Ocimene; 2, Linalool; 3, Geraniol; 4, Nerolidol; IS, internal standard. Abundance is scaled to the highest peak of all treatments per enzyme. Quantification of peaks in Table S6 and S7. B) Confirmation of identity of (*E*)-β-ocimene by comparison of mass spectra of (*E*)-β-ocimene produced in experiments and a standard. C) Pathway of how (*E*)-β-ocimene and linalool are formed from GPP.*

Functional experiments demonstrate the residual IDS activity of HMEL037106g1 and HMEL037108g1

While the production of terpenes can be tested by direct GC/MS analysis of the products of each experiment, this method will not detect isoprenyl diphosphates, potentially missing IDS activity if it is present. In order to test for IDS activity, we repeated the above experiments with DMAPP and IPP, GPP and IPP, and FPP and IPP, followed by treatment with alkaline phosphatase to hydrolyse the isoprenyl diphosphate products to their respective alcohols. These alcohols can then be detected by GC/MS analysis.

No further IDS activity was detected in either enzyme, apart from the residual IDS activity already determined above due to the trace amounts of terpenes produced from DMAPP and IPP. When either enzyme is provided with GPP, geraniol is produced, and when provided with FPP, large amounts of farnesol is produced, as expected from the dephosphorylation of the provided precursors, and this is seen in control conditions as well (Fig. S7, S8, Table S10, S11). As expected from the previous experiments, (*E*)-β-ocimene is also produced when HMEL037106g1 is provided with GPP, and linalool and nerolidol are produced when HMEL037108g1 is provided with GPP and FPP, respectively. Geranylgeraniol is not produced in any treatments, demonstrating that neither HMEL037106g1 nor HMEL037108g1 is a GGPPS, as suggested by their annotation (Fig. S7, S8, Table S10, S11). In summary, both HMEL037106g1 and HMEL037108g1 only exhibit residual IDS activity.

Evolutionary history of gene family containing Heliconius *TPSs*

Lineage-specific expansions of gene families are often correlated with functional diversification and the origin of novel biological functions (Lespinet et al. 2002). We therefore carried out a phylogenetic analysis of GGPPS in Lepidoptera to investigate whether gene duplication could have played a role in the evolution of the TPSs HMEL037106g1 and HMEL037108g1. Orthologs of the *H. melpomene* GGPPSs were identified in *H. cydno*, *Heliconius erato*, *Bicyclus anynana*, *Danaus plexippus*, *Papilo polytes*, *Pieris napi*, *Manduca sexta*, *Bombyx mori* and *Plutella xylostella* (Challis et al. 2016). Expansions of the GGPPS group of enzymes can be seen in *Heliconius* and in

Bicyclus, both groups in which terpenes form part of the pheromone blend (Bacquet et al. 2015) (Fig. S9).

To focus on the *Heliconius*-specific duplications, we made a phylogeny using the DNA sequence of transcripts from *H. melpomene*, *H. cydno* and *H. erato*. *Heliconius melpomene* and *H. cydno* belong to the same clade within *Heliconius*, with an estimated divergence time around 1.5 million years ago (Beltrán et al. 2002). *Heliconius erato* is more distantly related, belonging to a different *Heliconius* clade which diverged from the *H. melpomene*/*H. cydno* group around 10 million years ago (Kozak et al. 2015). Whilst (*E*)-β-ocimene is not found in the genitals of *H. cydno*, it is found in the genitals of *H. erato*, at around one tenth the amount of *H. melpomene* (Darragh et al. 2019b). We hypothesised that duplications between the *H. melpomene* and *H. erato* clades may have resulted in subfunctionalisation and a more efficient *H. melpomene* enzyme facilitating increased (*E*)-β-ocimene production. We found that both losses and gene duplications have occurred between the *H. melpomene* and *H. erato* clades, whilst gene copy number is conserved between closely-related *H. melpomene* and *H. cydno* (Fig. S10). The exact orthology between the *H. erato* and *H. melpomene*/*H. cydno* genes is unclear, but what is clear is that *H. melpomene*/*H. cydno* have more genes in this family than *H. erato* (Fig. S10), and that both clades have more genes than the ancestral lepidopteran state of one copy.

We also found evidence for the formation of pseudogenes following gene duplication. The amino acid sequences from translations of two genes in *H. melpomene*, *HMEL22305g1* and *HMEL037104g1*, do not contain complete functional protein domains. This is also seen for *Herato0606.241* in *H. erato*. Furthermore, more recent pseudogene formation could be seen in the *H. cydno* ortholog of *HMEL22306g3*, which contained multiple stop codons, despite exhibiting transcription (Fig. S3).

In order to determine the number of evolutionary origins of insect and plant TPSs we carried out a broader phylogenetic analysis, including other known insect and plant IDS and TPS proteins. Similar to the other insect TPSs described, *Heliconius* TPSs are not found within the same clade as plant ocimene synthases, representing an independent origin of ocimene synthesis in *Heliconius* and plants. Furthermore, the *Heliconius* TPSs do not group with known insect TPS enzymes in Hemiptera and Coleoptera (Fig. 6). Instead,

the *Heliconius* TPS enzymes group with GPP and GGPP synthases, rather than FPP synthases. The TPS enzymes of *Heliconius* are therefore of an independent evolutionary origin to other insect TPSs.

Comparison of the amino acid alignment of known insect TPSs with the *H. melpomene* enzymes (Fig. S11) demonstrated that residues previously identified as conserved in insect TPSs (Lancaster et al. 2018), were not found in the *H. melpomene* TPSs. No residues were shared between all insect TPSs (including *H. melpomene* TPS), which were not also shared with the *H. melpomene* GGPPS. This further indicates independent convergent evolution of TPS function in *H. melpomene*.

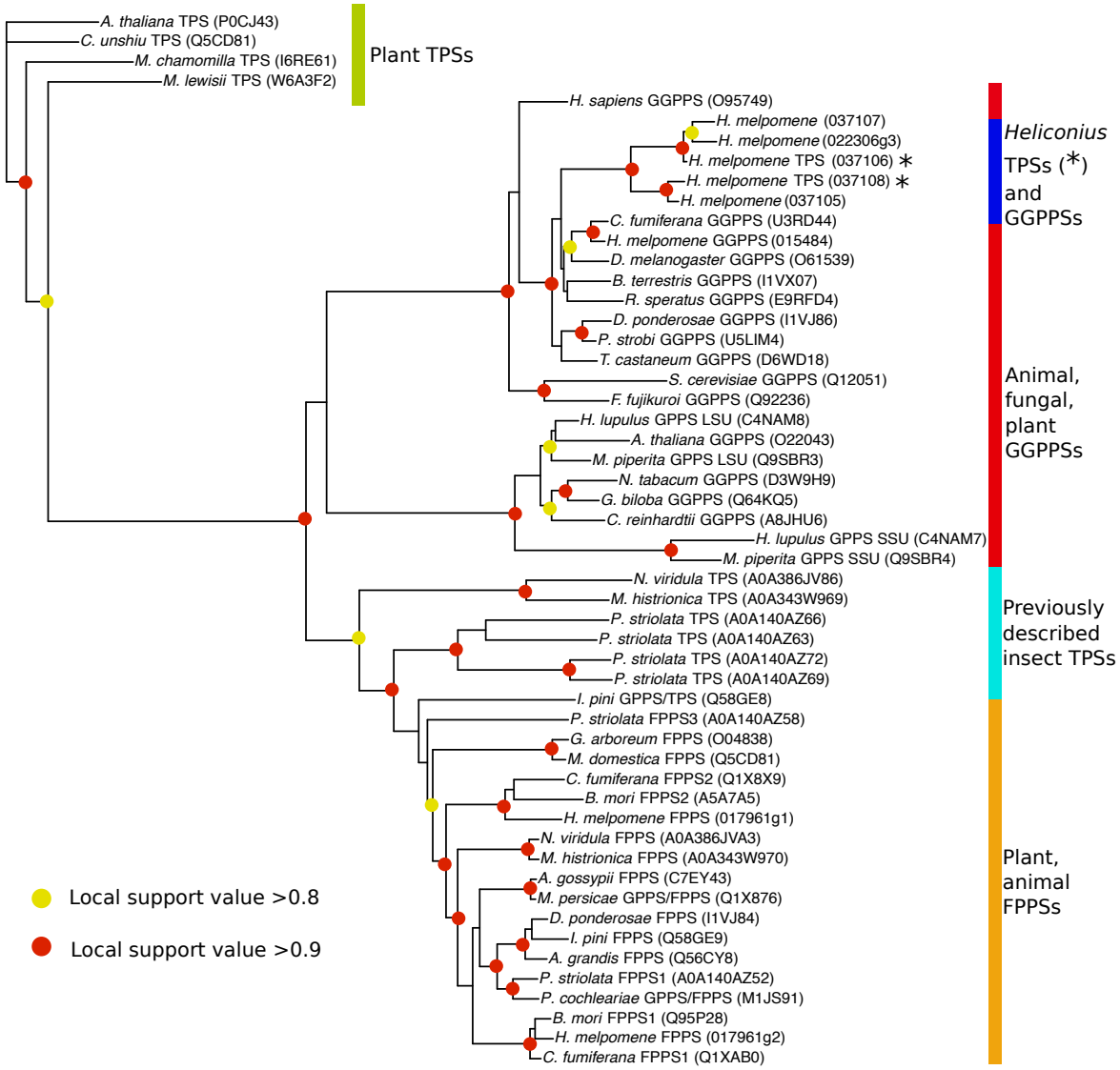


Figure 6: *Phylogenetic tree of GGPPS, FPPS, and TPS proteins of animals, fungi, and plants. The phylogeny was constructed by FastTree using the JTT (Jones-Taylor-Thornton) model of amino acid evolution. Local support values are illustrated. The tree was rooted with the ocimene synthase of* Citrus unshiu*. Full species names in Table S12.*

Discussion

Both plants and animals use terpenes as chemical signals, however, the terpene synthases that make them have been identified in only a few insect species. Ocimene is a common monoterpene and we have identified the first ocimene synthase in animals. We identify a region of the genome responsible for differences in ocimene production and discover a novel gene family within this region. We confirm ocimene synthase activity for one of these enzymes (HMEL037106g1), and terpene synthase activity in a closely related enzyme (HMEL037108g1). Neither of these genes are homologous to known plant TPSs and represent a novel TPS family in *Heliconius*. Furthermore, they are also very different from previously described insect TPSs. While the TPS enzymes of Hemiptera and Coleoptera are more closely related to FPPSs (Beran et al. 2016, 2019; Lancaster et al. 2018), the *H. melpomene* TPSs are more closely related to GGPPSs. We do not find shared amino acid changes with other insect TPSs, strongly suggesting that TPS activity in Lepidoptera has arisen independently. Overall, the origin of the (*E*)-β-ocimene synthase activity in *H. melpomene* represents an excellent example of chemical convergence via the independent evolution of new gene function.

Male *Heliconius melpomene* transfer ocimene to the female during mating. Within this context the biological function of HMEL037106g1 is clear, making the anti-aphrodisiac compound (*E*)-β-ocimene. However, the *in vivo* role of HMEL037108g1 is less clear. It is found within the QTL interval and also shows higher expression in *H. melpomene* males relative to *H. cydno*. This enzyme acts as a multifunctional linalool/nerolidol synthase, which have previously been described in plants (Zhu et al. 2014; Magnard et al. 2018). However, neither linalool or nerolidol are found in high amounts in the male abdomen (Darragh et al. 2019b). This apparent discrepancy may be due to the location or timing of expression *in vivo* (Beran et al. 2016). Another hypothesis is that *in vivo* GPP reacts with another substrate in the active site of HMEL037108g1, or that once linalool is produced it is metabolically channelled to another enzyme for further modification *in vivo* (Poshyvailo et al. 2017).This could also explain the lack of stereoselectivity in linalool formation. Further experiments will be required to determine if the other enzymes of this family not tested here exhibit TPS activity also.

Although we describe the first ocimene synthase in animals, ocimene synthases are likely to be found in other groups. Both *Bombus terrestris* and *Apis mellifera* use ocimene as a recruitment and larval pheromone, respectively (Granero et al. 2005; Maisonnasse et al. 2010). While the biosynthetic pathway is not known in these groups, a similar pathway to that proposed here has been suggested in *A. mellifera* (He et al. 2016). However, the existing data suggest that the loci responsible for ocimene synthesis are also likely to be independently evolved. Unlike *H. melpomene*, only one GGPPS is found in the *Apis* genome, whilst there are six FPPS genes, the result of lineage-specific duplications (Cheng et al. 2014). Although this needs to be confirmed by functional studies, based on the genomic patterns, we predict that convergence between Lepidoptera and Hymenoptera in the synthesis of ocimene also has an independent evolutionary origin.

Our findings also highlight the role that gene duplication plays in the evolution of new gene functions. Gene duplication is thought to be important for the evolution of new functions, as one gene copy can evolve a new function by a process called neofunctionalization (Conant and Wolfe 2008), often resulting in large gene families with related but different functions. These families follow a birth-and-death model of evolution, expanding and contracting through gene duplication, formation of pseudogenes, and gene deletion (Roelofs and Rooney 2003; Nei and Rooney 2005). Plant TPSs follow these dynamics, making up a large family formed of seven subfamilies, with lineage-specific expansions (Tholl 2006; Chen et al. 2011). Gene duplication followed by neofunctionalization has resulted in closely-related enzymes which can produce different compounds, in some cases due to subcellular localisation (Nagegowda et al. 2008).

Our data show a similar pattern of gene family diversification and suggest that gene duplication has facilitated the evolution of terpene synthesis in *Heliconius*. We uncover a lineage-specific expansion of GGPPSs in *Heliconius*. This novel gene expansion includes a number of pseudogenes, as well as two loci that possess TPS activity. A family of TPS genes has also been discovered in *Phyllotreta striolata,* where gene duplication is thought to have enabled functional diversification (Beran et al. 2016). Gene duplication can also facilitate enzyme specialisation by a process called subfunctionalization. In this case, an ancestrally multifunctional enzyme duplicates, resulting in two daughter copies

which split the ancestral functions, and can result in optimisation of these two functions (Conant and Wolfe 2008). Subfunctionalization might explain why neither *Heliconius* TPS shows significant IDS activity. In contrast to the multifunctional TPS/IDS enzyme from *I. pini* (Gilg et al. 2005, 2009), other insects have separate enzymes with IDS and TPS activity (Beran et al. 2016; Lancaster et al. 2018, 2019). One hypothesis is that an IDS enzyme initially gained TPS activity followed by gene duplication and subfunctionalization with enzymes specialised for different enzymatic steps. HMEL037106g1 is the first specialised monoterpene synthase described in animals.

We have identified a novel family of TPSs in *Heliconius* butterflies which is unrelated both to plant TPSs and to the few examples of previously described insect TPSs. We confirm that terpene synthesis has multiple independent origins in insects, which are themselves independent from the evolution of terpene synthesis in plants. Despite their independent evolution, insect TPSs show significant structural similarities, having evolved from IDS-like proteins. To understand how this diversity has arisen we need to identify the functional amino acid changes and relate structure to function, a nascent area of research for this group of enzymes (Abdallah and Quax 2017).

Materials and Methods

Analysis of biosynthetic pathway in H. melpomene

To identify genes involved in terpene biosynthesis we searched the *H. melpomene* genome (v2.5) on LepBase (Challis et al. 2016; Pinharanda et al. 2019) for genes in the mevalonate pathway and IDSs. *Drosophila melanogaster* protein sequences were obtained from FlyBase and used in BLAST searches (blastp) against all annotated proteins in the *H. melpomene* genome (Table S1) (Cheng et al. 2014; Thurmond et al. 2019). We used the BLAST interface on LepBase with default parameters (-evalue 1.0e-10 -num_alignments 25) (Priyam et al. 2015; Challis et al. 2016). We then searched these candidate orthologs against the annotated proteins of *D. melanogaster* using the BLAST interface on FlyBase to identify reciprocal best blast hits. We included in our results other hits with an e-value smaller than $1e^{-80}$.

Crossing for quantitative trait linkage mapping

To map the genetic basis of ocimene production we crossed *H. melpomene*, which produces (*E*)-β-ocimene, to *H. cydno*, a closely related species which does not. We crossed these two species to produce F1 offspring and backcross hybrids in both directions. Female F1s are sterile and so we mated male F1s to *H. cydno* and *H. melpomene* virgin stock females to create backcross families. Families created by backcrossing to *H. melpomene* had a phenotype similar to pure *H. melpomene*, suggesting the *H. melpomene* phenotype is dominant. While we used 265 individuals to create the linkage map, we focused on backcross families in the direction of *H. cydno*, where the (*E*)-β-ocimene phenotype segregates for the QTL mapping (Figure S1). We phenotyped and genotyped 114 individuals from 15 backcross families in the direction of *H. cydno*. Bodies were stored in dimethyl sulfoxide (DMSO) and stored at −20◦C for later DNA extraction.

Genotyping and linkage map construction

DNA extraction was carried out using Qiagen DNeasy kits (Qiagen). As previously described, individuals were genotyped either by RAD-sequencing (Davey et al. 2017; Byers et al. 2019; Merrill et al. 2019), or low-coverage whole genome sequencing using Nextera-based libraries (Picelli et al. 2014; Byers et al. 2019). A secondary purification using magnetic SpeedBeads™ (Sigma) was performed prior to Nextera-based library preparation. Libraries were prepared following a method based on Nextera DNA Library Prep (Illumina, Inc.) with purified Tn5 transposase (Picelli et al. 2014). PCR extension with an i7-index primer (N701–N783) and the N501 i5-index primer was performed to barcode the samples. Library purification and size selection was done using the same beads as above. Pooled libraries were sequenced by BGI (China) using HiSeq X Ten (Illumina).

Linkage mapping was conducted following (Byers et al., 2019), using standard Lep-MAP3(LM3) pipeline (Rastas 2017). Briefly, fastq files were mapped to the *H. melpomene* reference genome using BWA MEM (Li 2013). Sorted bams were then created using SAMtools and genotype likelihoods constructed (Li et al. 2009). The pedigree of individuals was checked and corrected using IBD (identity-by-descent) and the sex checked using coverage on the Z chromosomes by SAMtools depth. A random subset of

25% of markers were used for subsequent steps. Linkage groups and marker orders were constructed based on the *H. melpomene* genome and checked with grandparental data.

The map constructed contained 447,820 markers. We reduced markers by a factor of five evenly across the genome resulting in 89,564 markers with no missing data to facilitate computation. We log-transformed amounts of ocimene produced to conform more closely to normality. Statistical analysis was carried out using R/qtl (Broman et al. 2003). We carried out standard interval mapping using the *scanone* function with a non-parametric model, an extension of the Kruskal-Wallis test statistic. The analysis method for this model is similar to Haley-Knott regression (Haley and Knott 1992). We used permutation testing with 1000 permutations to determine the genome-wide LOD significance threshold. To obtain confidence intervals for QTL peaks we used the function *bayesint.* Phenotype data, pedigree, linkage map and R script is available from OSF (https://osf.io/3z9tg/?view_only=63ba7c0767a84d8eb907fbf599df062f). Sequencing data used for the linkage maps is available from ENA project ERP018627 (Byers et al. 2019).

*Phenotyping of (*E*)-β-ocimene production*

Chemical extractions were carried out on genital tissue of mature (7-14 days post-eclosion) male individuals of *H. melpomene, H. cydno,* and hybrids (for details of butterfly stocks please see SI Material and Methods). Genitals were removed using forceps and soaked, immediately after dissection, in 200µl of dichloromethane containing 200 ng of 2-tetradecyl acetate (internal standard) in 2ml glass vials with polytetrafluoroethylene-coated caps (Agilent, Santa Clara, California). After one hour, the solvent was transferred to a new vial and stored at −20 ◦C until analysis by gas chromatography-mass spectrometry (GC-MS). For details of GC-MS analysis please see SI Materials and Methods.

RNA sequencing analysis

Gene expression analyses were performed using already published RNA-seq data from heads and abdomens of *H. melpomene* and *H. cydno* from GenBank BioProject PRJNA283415 (Walters et al. 2015). Although it would be possible to map the *H.cydno*

RNA-seq reads to *H. melpomene* due to high genome sequence similarity, that might lead to biases associated with reads carrying *H. cydno* specific alleles. To accurately quantify gene expression in *H. cydno* we generated an assembly and annotation of the H. cydno genome using sequencing data available from ENA study ERP009507 (Davey et al. 2017). We then manually identified the *H. cydno* orthologs of our seven candidate genes (Table S2) and curated the annotation to make it compatible with RNA-seq analysis software (for details of the assembly and annotation please see SI Materials and Methods). We performed quality control and low quality base and adapter trimming on the RNA-seq data using TrimGalore! (Martin 2011). We then mapped the reads to the *H.melpomene* genome v2.5 (Davey et al. 2016) and our newly assembled *H.cydno* genome using STAR (Dobin et al. 2013). *featureCounts* (Liao et al. 2014) was used to produce read counts that were normalised by library size with TMM (trimmed mean of M values) normalisation (Robinson and Oshlack 2010) using the edgeR package in R (Robinson et al. 2010). To test for differences in expression of our candidate genes, we used the *voom* function from the limma package in R (Law et al. 2014), which fits a linear model for each gene by modelling the mean-variance relationship with precision weights.

To test for male abdomen-biased expression within *H. melpomene* we included two fixed effects, sex and tissue, as well as including individual as a random effect (expression ~ sex + tissue + sex*tissue + (1|individual)). We were looking for genes with a significant interaction between sex and tissue, showing higher expression in male abdomens. To test for differences in expression between *H. melpomene* and *H. cydno* abdomens we included two fixed effects, sex and species, as well as an interaction term (expression ~ sex + species + species*tissue). We were interested in finding differences in the extent of sex-bias between species, again detected by a significant interaction term with higher expression in *H. melpomene* male abdomens.

P-values were corrected for multiple testing using the Benjamini-Hochberg procedure for all genes in the genome-wide count matrix (17902 for *H. melpomene*). For the interspecific comparison we identified genome wide orthologs from the annotation and produced a gene count matrix including both species. The ortholog list was limited to genes that had only one ortholog in each species (11571 genes). Scripts are available from OSF (https://osf.io/3z9tg/?view_only=63ba7c0767a84d8eb907fbf599df062f).

In vitro *expression and enzymatic assays*

For more details of the expression and enzymatic assays please see SI Materials and Methods. Briefly, cDNA libraries were synthesised from RNA extracted from male abdominal tissue of *H. melpomene*. Genes of interest were amplified by PCR with gene-specific primers (Table S5), purified, sequenced for confirmation, ligated into expression plasmids, and transformed into competent *Escherichia coli* cells. Cell cultures were grown to an OD_{600} of 0.5 were induced with 1mM IPTG and cultivated for a further two hours before collection by centrifugation. Cells were resuspended in assay buffer and sonicated.

TPS and IDS activity was assayed using the soluble fraction of the cell lysate. We added either DMPP and IPP, GPP and IPP, FPP and IPP, or GPP alone. We also tested for enzymatic activity with (*R*)-linalool and (*S*)-linalool. For TPS activity assays, reactions were immediately stopped on ice and extracted with hexane. For IDS activity assays, reaction mixtures were incubated with alkaline phosphatase to hydrolyse the pyrophosphates before hexane extraction. Prior to analysis by GC-MS, an internal standard was added and samples concentrated. Products were compared to control experiments without protein expression. For details of GC-MS analysis please see SI Materials and Methods.

Phylogenetic analysis

To identify orthologs of the GGPPS in other Lepidoptera we searched protein sequences from *H. melpomene* version 2.5 (Davey et al. 2016, 2017) against the genomes of *H. erato demophoon* (v1), *Bicyclus anynana* (v1x2), *Danaus plexippus* (v3), *Papilo polytes* (ppol1), *Pieris napi* (pnv1x1), *Manduca sexta* (msex1), *Bombyx mori* (asm15162v1), and *Plutella xylostella* (pacbiov1), using the BLAST interface (tblastn) on LepBase (Priyam et al. 2015; Challis et al. 2016). We also included the previously identified orthologs from the *H. cydno* genome (Table S2). To check that the predicted orthologs contained functional protein domains we used the NCBI conserved domain search with default parameters (Marchler-Bauer et al. 2015). We deleted any proteins found without complete functional domains, including a gene from *H. erato*, *Herato0606.241*. We also did not include the *H. cydno* ortholog of *HMEL22306g3* in the

protein tree, as despite showing transcription (Fig. S3), there were multiple stop codons within the coding region.

To focus on the *Heliconius*-specific duplications, we downloaded the transcript sequences for the *H. melpomene* and *H. erato* proteins from LepBase and exported transcripts for predicted genes in Apollo for *H. cydno*. (Table S2). We used gene *Herato0606.245* (GGPPS, shows high similarity to the GGPPS of the moth *Choristoneura fumiferana*) to root the tree.

To investigate the evolutionary relationship of the *Heliconius* GGPPS we carried out a broader phylogenetic analysis with other known insect and plant IDS and TPS proteins. Protein sequences for these additional enzymes were downloaded from Uniprot (The UniProt Consortium 2019). *Heliconius* protein sequences were obtained as described above. We used an ocimene synthase enzyme from *Citrus unshiu* to root the tree.

We aligned amino acid or DNA sequences using Clustal Omega on the EMBL-EBI interface (Madeira et al. 2019). Alignments were visualised using BoxShade (https://embnet.vital-it.ch/software/BOX_form.html). Phylogenies were inferred using FastTree, a tool for creating approximately-maximum-likelihood phylogenetic trees, with default parameters (Price et al. 2009, 2010). These phylogenies were plotted using the package *ape* and *evobiR* in R version 3.5.2. (Adams 2015; Paradis and Schliep 2018; R Core Team 2018). To ensure correct placement of support values when re-rooting trees we checked phylogenies using Dendroscope (Huson and Scornavacca 2012; Czech et al. 2017). Phylogenies and R code are available from OSF (https://osf.io/3z9tg/?view_only=63ba7c0767a84d8eb907fbf599df062f).

Supplementary Information

Butterfly stocks

Outbred stocks of *Heliconius melpomene rosina* and *H. cydno chioneus* were established from wild individuals collected in Gamboa (9˚7.4' N, 79˚42.2' W, elevation 60 m) in the nearby Soberania National Park, San Lorenzo National Park (9°17'N, 79°58'W; elevation 130 m), and in Altos de Campana National Park (8°69' N, 79°92' W; elevation 900 m). Stocks were maintained in insectaries at the Smithsonian Tropical Research Institute (STRI) facilities in Gamboa, Panama. Individuals for this study were reared under ambient conditions between January 2016 and January 2018. Larvae were reared on *Passiflora platyloba*. Adult male butterflies were kept in cages with other males and provided with approximately 20% sucrose solution with access to at least one of *Psychotria poeppigiana*, *Gurania eriantha*, *Psiguiria triphylla* and *Psiguria warscewiczii* as pollen sources.

GC/MS analysis

DMAPP (90%), GPP (95%) and IPP (95%) were purchased from Sigma. FPP was purchased from VWR (98%). (*R*)-Linalool was purchased from Merck (95%). (*S*)-Linalool was purchased from Sigma as coriander oil and purified through column chromatography.

Extracts from adult butterflies, and samples from *in vitro* experiments were analysed by gas chromatography/mass spectrometry (GC/MS) using an Agilent (model 5977/5975) mass-selective detector connected to an Agilent GC (model 7890B/7890A) with electron impact ionisation (70eV). This instrument was equipped with an Agilent ALS 7693 autosampler and an HP-5MS fused silica capillary column (Agilent) (length 30 m, inner diameter 0.25 mm, film thickness 0.25 µm). Injection was performed in splitless mode (injector temperature 250°C) with helium as the carrier gas (constant flow of 1.2 ml/min). The temperature programme started at 50°C for 5 min, and then rose at a rate of 5°C /min to 320°C, before being held at 320°C for 5 min. The compounds were identified through comparison with retention time and mass spectra of standard samples.

Chiral analysis of linalool was performed in an Agilent 7820A gas chromatograph equipped with flame ionization detector (FID), using a chiral column Beta DEX 225 (length 30m, inner diameter 0.25mm, film thickness 0.25µm, Supelco). The oven program started at 50°C for 1 minute, followed by increasing the temperature at a rate of 3°C/min until 210°C, keeping this temperature for 5 minutes. Samples were injected in splitless mode and flow of 1.65 mL/min. The peak areas were used to calculate the percentage of each stereoisomer.

Heliconius cydno guided assembly and annotation transfer

H. cydno and *H. melpomene* had their most recent common ancestor 1.5 million years ago and their absolute divergence is roughly 3% (dxy ~0.03) (Kozak et al. 2015; Martin et al. 2016). Due to this high degree of similarity, it is possible to map *H. cydno* RNA-seq reads to the *H. melpomene* genome. However, we wanted to accurately quantify gene expression in existing *H. cydno* samples (GenBank BioProject PRJNA283415 (Walters et al. 2015)) by reducing potential biases associated with RNA-seq reads carrying H*. cydno* specific alleles. RNA-seq reads from *H. cydno* with such variants have a lower probability to map correctly to the existing *H. melpomene* reference. This biases quantification and increases false positive rates, documented extensively, specifically in the context of allele-specific expression studies. (Degner et al. 2009).

A *H. cydno* trio (mother, father and progeny) was previously Illumina sequenced (ENA study ERP009507) and assembled into maternal and paternal genomes with trio-sga (Malinsky et al. 2016). The paternal genome had 34,566 scaffolds, a total size of 270,339,622 bp and a scaffold N50 of 25,716 bp, with 551 kb of gaps (paternal trio fasta file available from OSF (https://osf.io/3z9tg/?view_only=63ba7c0767a84d8eb907fbf599df062f)). To improve gene contiguity we used the progressiveCactus algorithm (v3) to align the *H. cydno* paternal haplotypic assembly to the chromosomal version of the *H. melpomene* genome (v2.5, Paten et al. 2011a,b; Davey et al. 2016). The HAL database created by progressiveCactus was loaded to Ragout (v1.2, Kolmogorov et al. 2014) to produce the final *H. cydno* reference-guided assembly (*H. cydno* reference fasta file; ordering information and unplaced scaffolds available from OSF

(https://osf.io/3z9tg/?view_only=63ba7c0767a84d8eb907fbf599df062f). The *H. cydno* guided assembly has 58 scaffolds, a total size of 261,056,210 bp, a scaffold N50 of 13,724,118 bp, and 8.3 Mb of gaps.

We then transferred the *H. melpomene* annotation (v2.5) to the *H. cydno* assembly. We used EMBOSS Seqret (v6.6.0.0) to convert the *H. melpomene* annotation file to the embl format (Rice et al. 2000) and we used RATT to transfer the *H. melpomene* annotation (reference) to the guided *H. cydno* genome (query). RATT is part of PAGIT, a post-assembly genome-improvement toolkit (v1.0) (Swain et al. 2012). We searched for synteny between the reference and the query using MUMmer (v4.0) and detected possible errors such as start and stop codons or frameshift mutations (Kurtz et al. 2004). After correcting such errors with the RATT pipeline the annotation transfer to *H. cydno* was complete (Otto et al. 2011).

To ensure that our genes of interest from *H. melpomene* (those identified in Table S1) were correctly annotated we manually curated these genes in the *H. cydno* annotation. To find orthologs in *H. cydno* we used the BLAT function in Apollo to search for *H. melpomene* exons (Kent 2002; Lewis et al. 2002). We checked the gene models for splice sites and start and stop codons. The curated gene models were then exported from Apollo and manually included in the *H. cydno* annotation. We then subset the annotation to include only exons, because CDS sequences had not been properly annotated (Updatedannotation.gff). We then converted it to gtf file format using the gffread function of Cufflinks (Hcyd1.0_annotV2.gtf) (Trapnell et al. 2012) and filtered out exons longer than 30,000 base pairs (Hcyd1.0_annotV2.gtf; gtf_modify_Hcyd_annotV2.R). We finally used the gtf_modify_Hcyd_annotV3.R script to include unique *H. cydno* gene-ids (Hcyd1.0_annotV3.gtf).

In vitro *expression and enzymatic assays*

RNA extraction from male abdominal tissue of *H. melpomene* was carried out following a standard TRIzol protocol (Invitrogen) and cDNA synthesised using 5x iScript Reaction Mix (Bio-Rad). Following the protocol from *Champion™ pET101 Directional TOPO™ Expression Kit* (Invitrogen), we amplified the full-length transcript of genes of

interest from the cDNA by PCR using Q5 High-Fidelity 2x Master Mix (Biolabs), with gene-specific primers (Table S5). The primers were designed for full-length transcript amplification. The PCR products were purified using a MiniElute PCR purification kit (Qiagen) and then sequenced to confirm identity (Table S5). Following sequencing, the PCR products were ligated into the expression vector pET101/D-TOPO® and transformed into One Shot® TOP10 Chemically Competent *Escherichia coli* cells. Plasmids were extracted from cultures of successful colonies using the QIAprep Spin Miniprep Kit (Qiagen) and sequenced again to confirm correct ligation in the vector using the T7 and T7-reverse primers (Table S5).

Plasmids containing the genes of interest in the correct orientation were transformed into *Escherichia coli* strain BL21 Star™(DE3) for expression. Cell cultures were grown to an OD_{600} of 0.5 and induced with 1mM IPTG. After induction, the cells were cultivated for a further two hours at 37°C and 250rpm, before collection by centrifugation for 15 minutes at 6000xg at 4°C. Expression of protein was verified using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE). Pellets were resuspended in chilled extraction buffer (25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid pH 7.5, 1 mM MnCl2, 100 mM KCl, 3 mM dithiothreitol, 10% glycerol, protease inhibitor cocktail (Sigma)) and disrupted by sonication. Cell lysates were then centrifuged for 10 minutes at 9000xg at 4°C and the supernatant (containing the soluble part of the cell lysate) retained.

TPS and IDS activity was assayed using the soluble fraction of the cell lysate. Protein concentration was estimated using a Qubit® Protein Assay Kit (Invitrogen). 80-100ng of protein was added to each reaction in a total volume of 300µl. We added different precursors from different steps in the pathway (Fig. 1) to characterise enzymatic activity. Experiments were incubated at 30°C for two hours at 200rpm.

Firstly, we added DMAPP and IPP (100µM each), the two building blocks at the beginning of the pathway. To form a terpene from these compounds, they first need to be combined to form GPP, which can then be converted to a terpene by TPS activity. If the enzyme is a multifunctional GPPS/TPS, as in *Ips pini*, monoterpenes should be formed from DMAPP and IPP, via the production of GPP. Furthermore, if FPPS or GGPPS activity is

present, FPP or GGPP could be formed from DMAPP and IPP, as well as sesquiterpene or diterpenes if sesquiterpene or diterpene synthase activity is exhibited. We then carried out assays with GPP (100µM) and IPP (50µM). If the enzyme solely exhibits monoterpene synthase activity, the monoterpene could only be formed from GPP directly and not from DMAPP and IPP. Furthermore, the enzyme could be an FPPS or GGPPS, and could therefore produce FPP or GGPP from GPP and IPP. FPP and GGPP could be converted to sesquiterpenes or diterpenes if sesquiterpene or diterpene synthase activity is exhibited. We also tested with GPP alone (100µM) to test for monoterpene synthase activity directly. Finally, we carried out assays with FPP (100µM) and IPP (50µM). If the enzyme is a GGPPS as annotated, it should form GGPP from FPP and IPP, as well as potentially converting GGPP to diterpenes. This is also a test for sesquiterpene synthase activity, as sesquiterpenes should be formed from FPP if the enzyme is a sesquiterpene synthase. We also tested for enzymatic activity with (*R*)-linalool and (*S*)-linalool (100µM).

To test for IDS activity, we repeated the above experiments with DMAPP and IPP, GPP and IPP, and FPP and IPP, followed by treatment with alkaline phosphatase to hydrolyse the isoprenyl diphosphate products to their respective alcohols. These alcohols can then be detected by GC/MS analysis.

Dephosphorylation of GPP produces the monoterpene alcohol geraniol, whilst dephosphorylation of FPP produces the sesquiterpene alcohol farnesol. Our expectations for controls, without IDS or TPS enzymatic activity, is to find geraniol when GPP is provided, and farnesol when FPP is provided. Linalool is a monoterpene alcohol which is an isomer of geraniol, and nerolidol is a sesquiterpene alcohol which is an isomer of farnesol. Furthermore, geranylgeraniol is a diterpene alcohol derived from the dephosphorylation of GGPP. If an enzyme is exhibiting IDS activity, we expect it to be able to catalyse the condensation of IPP and the other precursor provided, DMAPP, GPP, or FPP, to form larger molecules. Therefore, when provided with DMAPP and IPP, we expect to find either monoterpene or sesquiterpene alcohols, derived from GPP or FPP. When provided with GPP and IPP, we expect sesquiterpene alcohols derived from FPP. When provided with FPP and IPP, we expect larger diterpene alcohols, such as geranylgeraniol, to form via the formation of GGPP.

For TPS activity assays, reactions were stopped on ice and overlaid with 250µl hexane and left at 25°C overnight. The hexane layer was then transferred to a new vial and stored at -80°C. For IDS activity assays, following incubation with the precursors, 20 units of alkaline phosphatase (Sigma) in alkaline phosphatase buffer was added to each reaction mixture and incubated at 30°C for four hours at 200rpm. Following this, 250µl hexane was added and left at 25°C overnight. The hexane layer was then transferred to a new vial and stored at -80°C. Prior to analysis by GC/MS, 20 µL of a solution of 2-acetoxytetradecane in hexane (10 µg/mL) was added as an internal standard, and samples concentrated to a volume of approximately 30 µL. Products were compared to control experiments without protein expression.

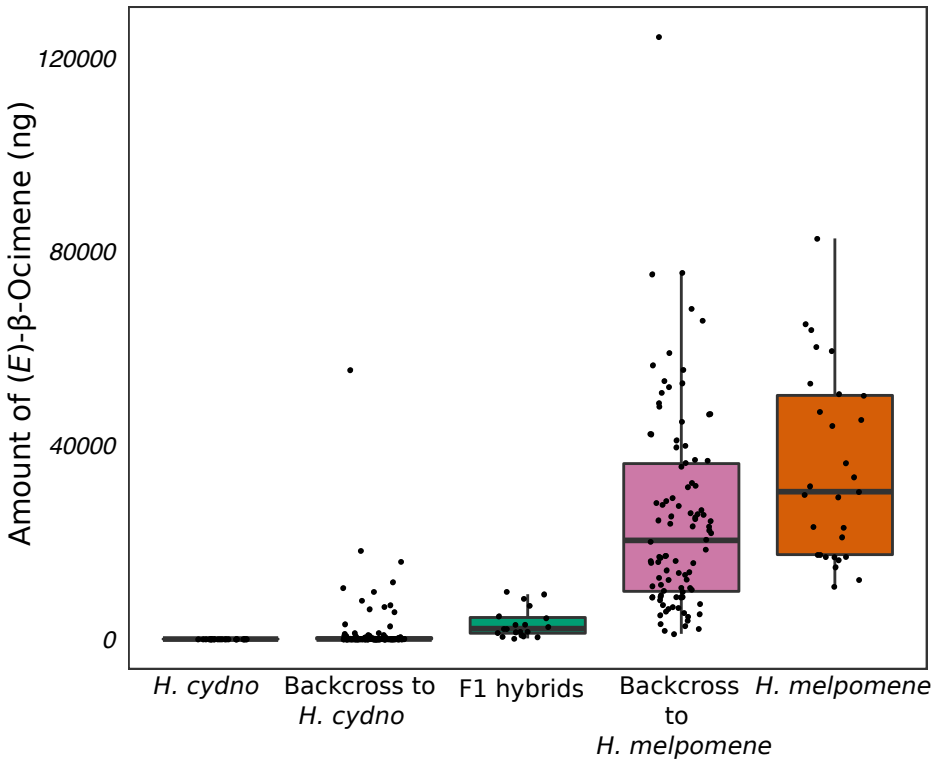


Figure S1: *Amount of (*E*)-β-ocimene (ng) in both pure parental species, F1 hybrids, and backcrosses in both directions. The phenotype segregates in backcrosses to* H. cydno *and, therefore, we focused on this cross direction.*

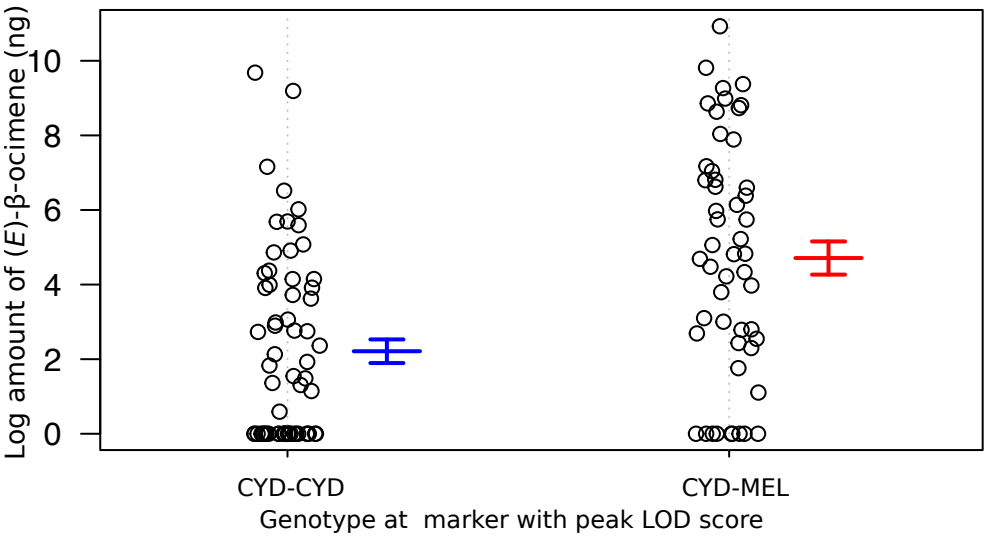


Figure S2: *Effect plot for QTL peak on chromosome 6. Log amount of (*E*)-β-ocimene produced by each genotype at the marker with the highest LOD score. Individuals homozygous for* H. cydno *alleles produce less (*E*)-β-ocimene than heterozygotes with a* H. melpomene *allele.*

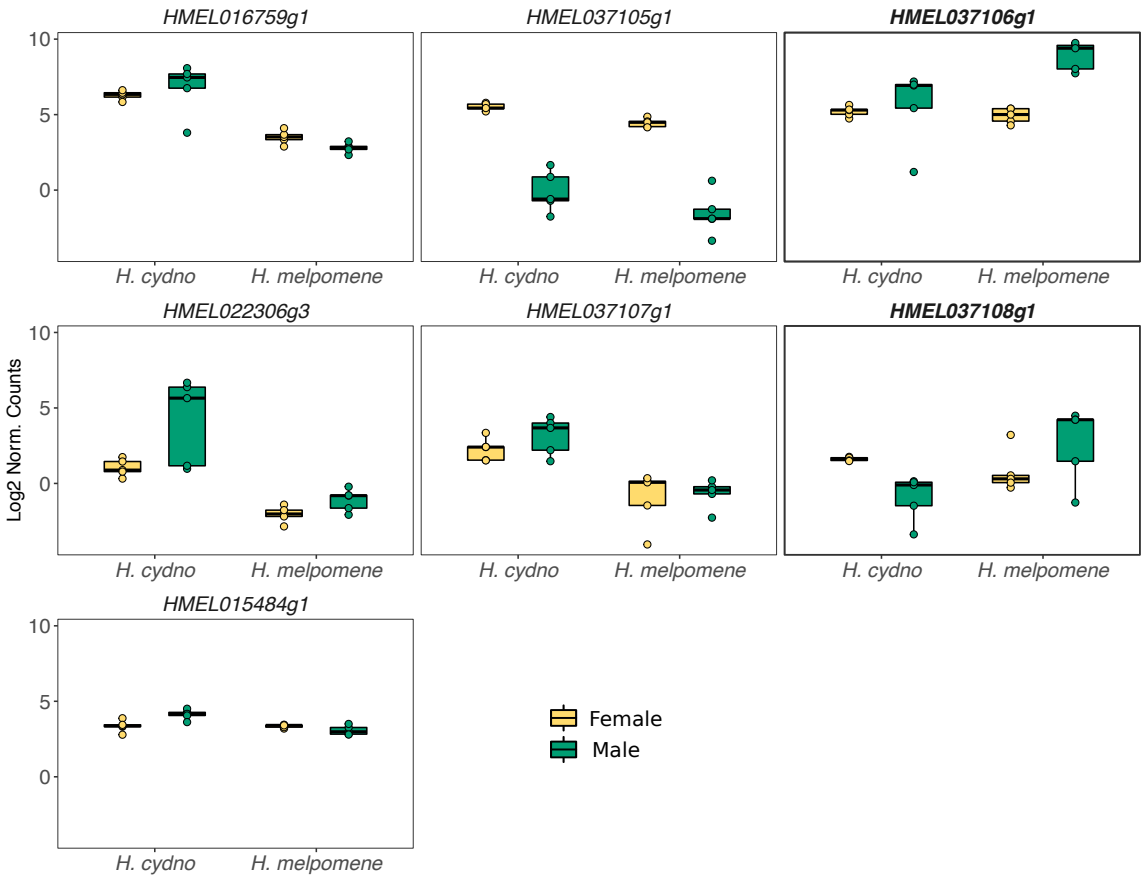


Figure S3: *Log2-expression of the candidate genes in* H. melpomene *and* H. cydno *abdomens in males and females. females. Both HMEL037106g1 and HMEL037108g1 (highlighted in bold) show greater male-biased expression in* H. melpomene *than* H. cydno*. Full model statistics in Table S4. N=5 for each boxplot. Gene expression is given in the log2 of the normalised counts using TMM (trimmed mean of M values) normalisation.*

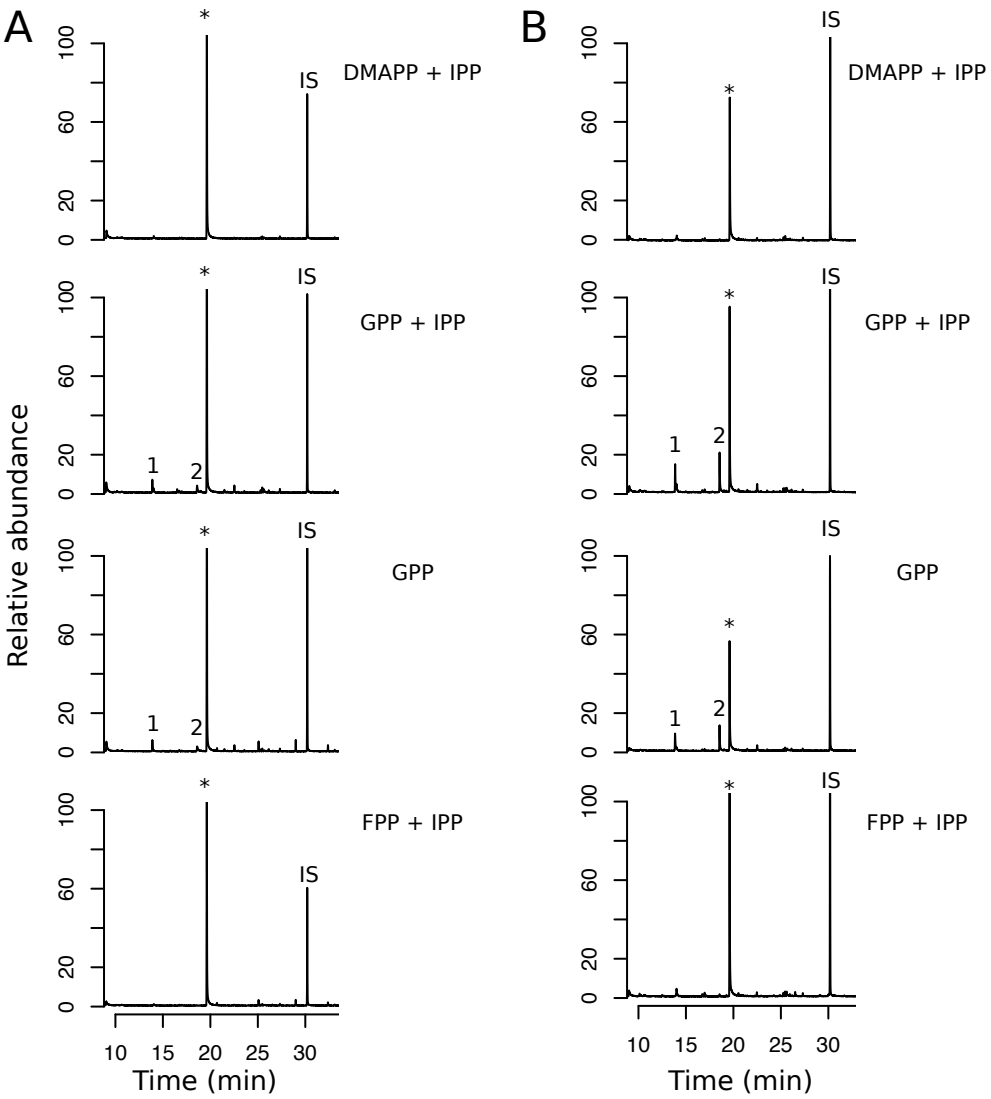


Figure S4: *Control experiments (protein expression uninduced) for the functional characterisation of TPS activity of A) HMEL037106g1 and B) HMEL037108g1 from* H. melpomene*. Total ion chromatograms of products in the presence of different precursor compounds. (*E*)-β-Ocimene is not produced in any treatments. Linalool and geraniol are produced in small amounts in both, likely due to endogenous bacterial activity. 1, Linalool; 2, Geraniol; *, contaminant from medium; IS, internal standard. Abundance is scaled to the highest peak of all panels per enzyme. Quantification of peaks in Table S6 and S7.*

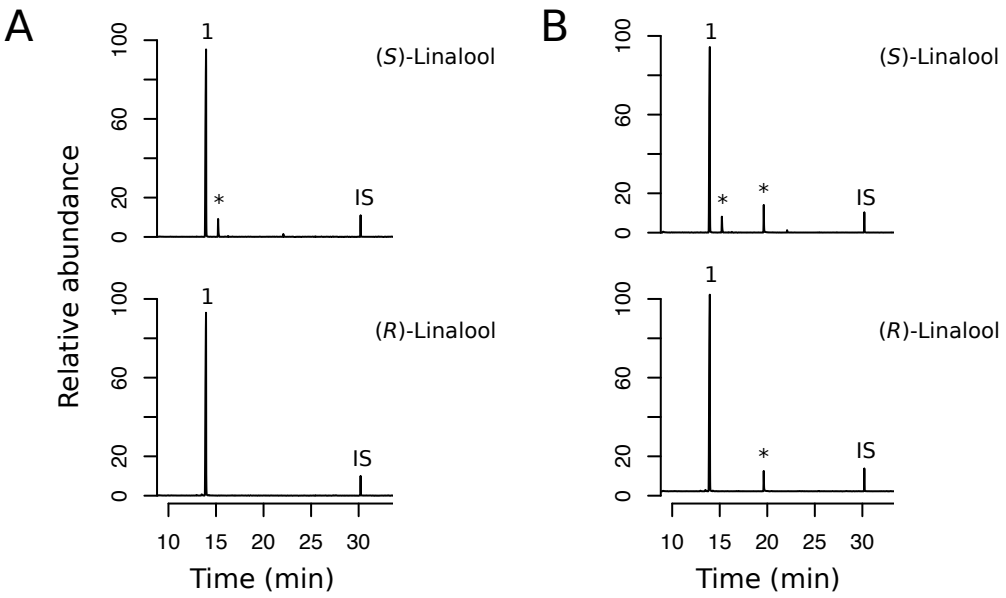


Figure S5: *Linalool is not metabolized into ocimene by HMEL037106g1. A) Total ion chromatograms of enzymatic products in the presence of different linalool stereoisomers. No enzymatic activity is detected. B) Total ion chromatograms of control experiments (protein expression not induced) in the presence of different Linalool stereoisomers. Again, as expected, no enzymatic activity is detected. 1, Linalool; *, contaminants from medium; IS, internal standard. Abundance is scaled to the highest peak of all panels. Quantification of peaks in Table S8.*

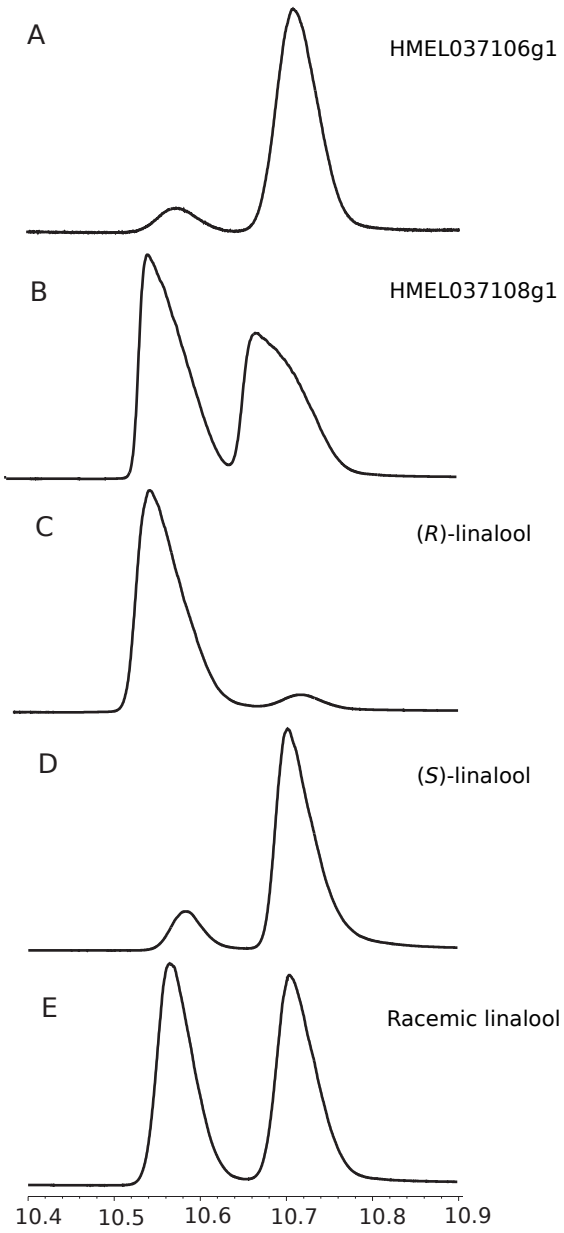


Figure S6: *Chiral analysis of linalool produced by HMEL037106g1 and HMEL037108g1. A) Linalool produced in experiments with HMEL037106g1 is mainly (*S*)-linalool (ratio 97:3,* S:R*), B) linalool produced in experiments with HMEL037108g1 is a racemic mixture (ratio 54:56,* S:R*), C) (*R*)-linalool, D) (*S*)-linalool, E) Racemic linalool mixture.*

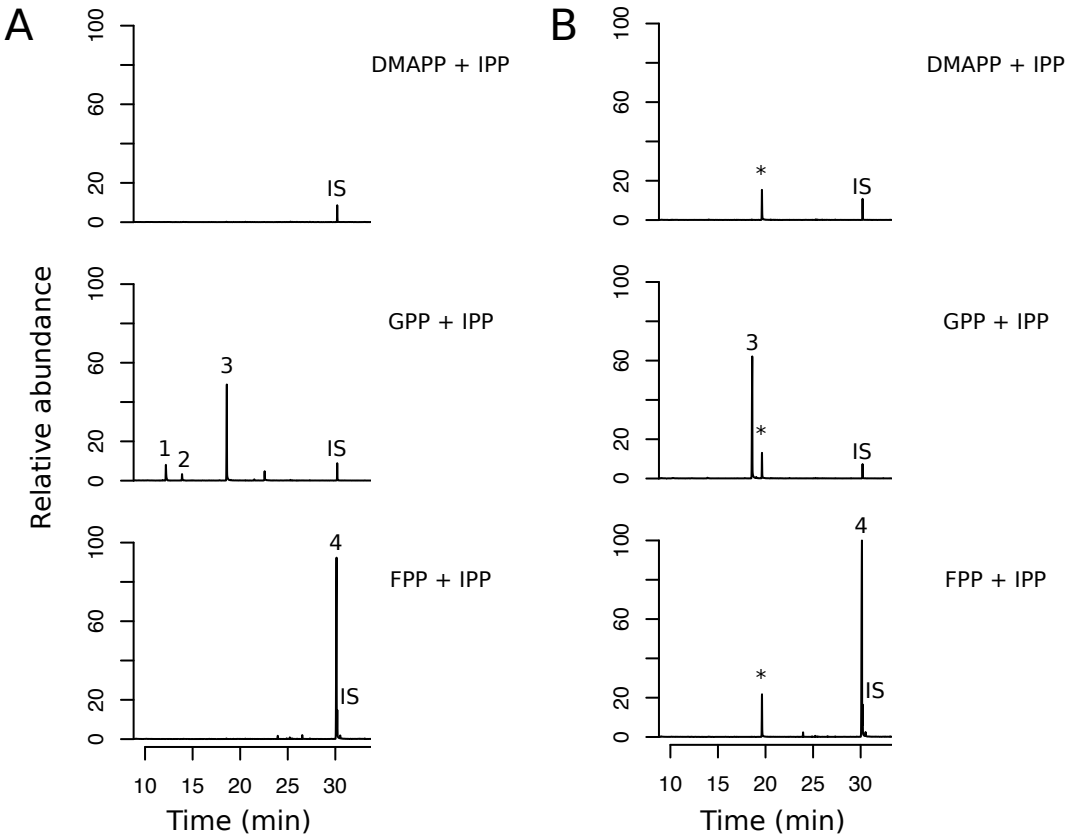


Figure S7: *Functional characterisation of IDS activity of HMEL037106g1 from* H. melpomene*. A) Total ion chromatograms of enzymatic products in the presence of different precursor compounds, following treatment by alkaline phosphatase. GPP is dephosphorylated to produce geraniol, and FPP to produce farnesol, demonstrating that the main function of HMEL037106g1 is not as an IDS. B) Total ion chromatograms of control experiments (protein expression not induced) in the presence of different precursor compounds, following treatment by alkaline phosphatase. As expected, GPP is dephosphorylated to geraniol, and FPP to farnesol. 1,* (E)-β-*Ocimene; 2, Linalool; 3, Geraniol; 4, Farnesol; *, contaminant from medium; IS, internal standard. Abundance is scaled to the highest peak of all panels. Quantification of peaks in Table S10.*

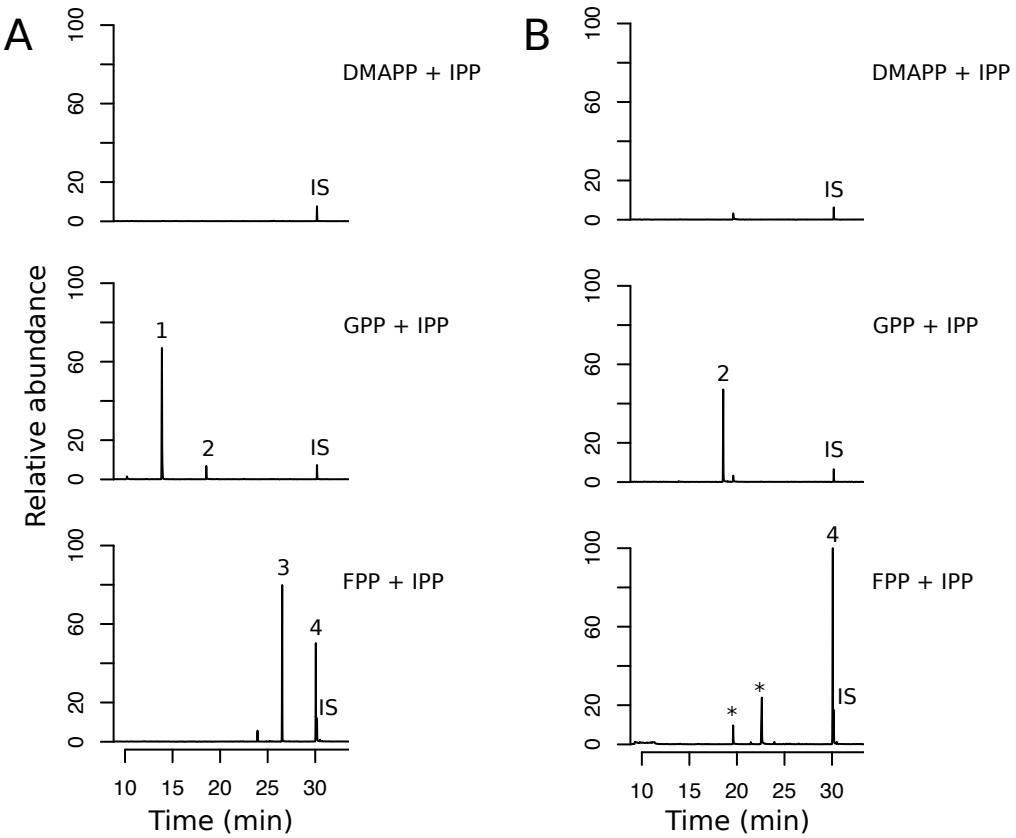


Figure S8: *Functional characterisation of IDS activity of HMEL037108g1 from* H. melpomene*. A) Total ion chromatograms of enzymatic products in the presence of different precursor compounds, following treatment by alkaline phosphatase. As in Figure S6, GPP is converted to linalool and FPP to nerolidol, with remaining GPP dephosphorylated to geraniol, and FPP to farnesol. HMEL037108g1 is acting as a mono- and sesquitepene synthase, not an IDS. B) Total ion chromatograms of control experiments (protein expression not induced) in the presence of different precursor compounds, following treatment by alkaline phosphatase. GPP is dephosphorylated to geraniol and FPP to farnesol. 1, Linalool; 2, Geraniol; 3, Nerolidol; 4, Farnesol; *, contaminant from medium; IS, internal standard. Abundance is scaled to the highest peak of all panels. Quantification of peaks in Table S11.*

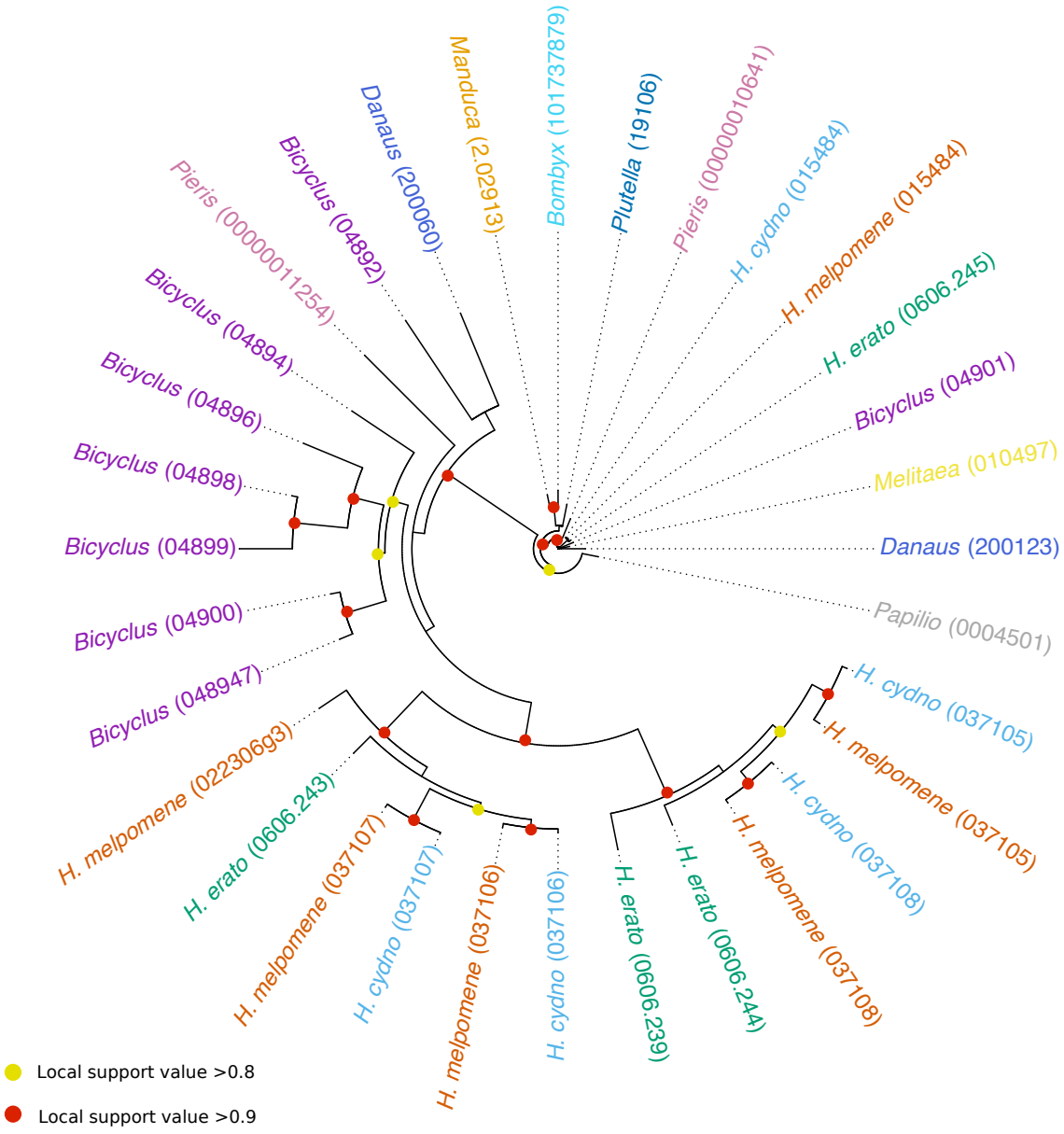


Figure S9: *Unrooted phylogenetic tree showing the relationships between protein sequences of GGPPSs in Lepidoptera. The tree was obtained from FastTree, a tool for creating approximately-maximum-likelihood trees, using the JTT (Jones-Taylor-Thornton) model of amino acid evolution. Local support values are illustrated.*

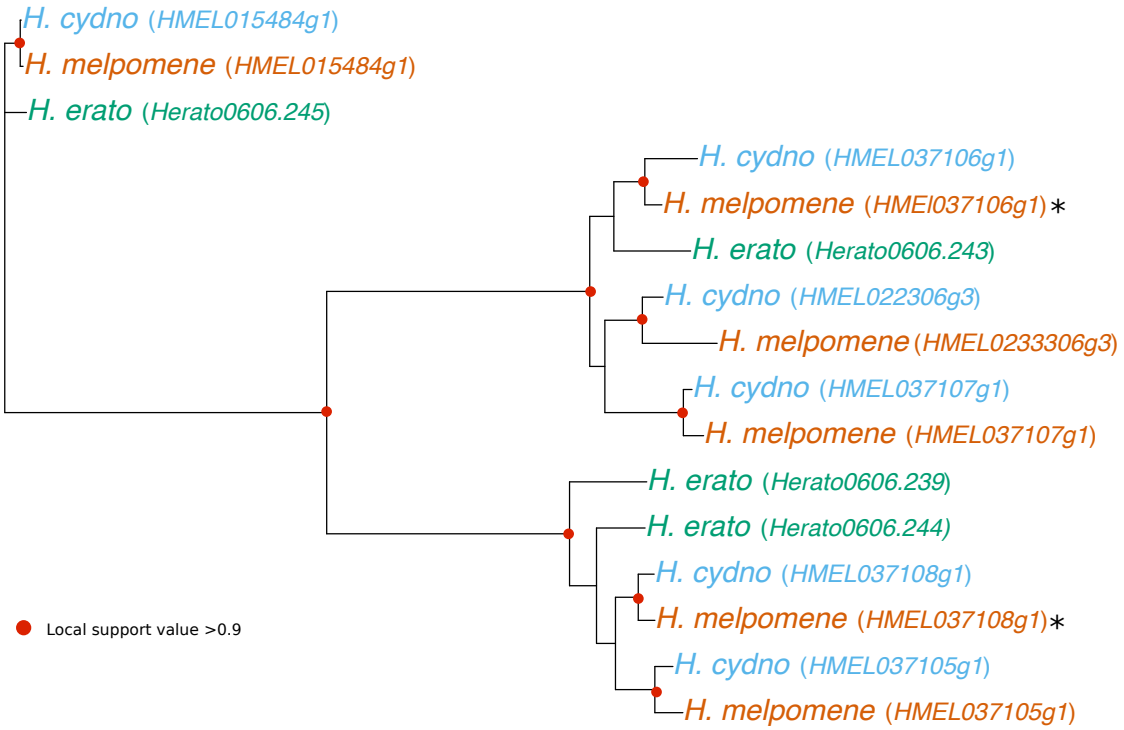
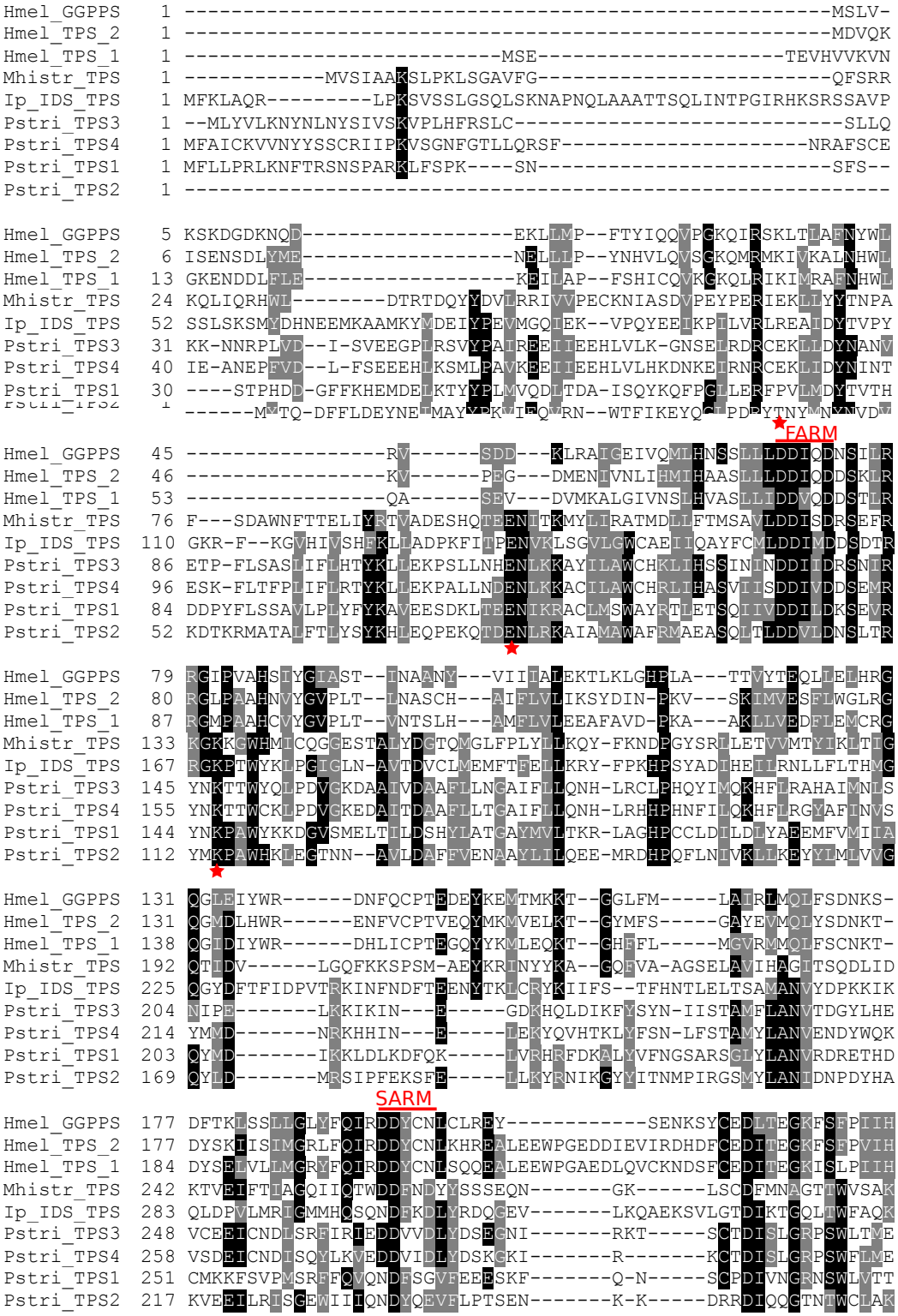


Figure S10: *Phylogenetic tree of genes annotated as GGPPSs in* Heliconius melpomene*,* H. cydno*, and* H. erato*, including* HMEL037106g1 *and* HMEL037108g1 (*) *which act as TPSs. The phylogeny was constructed in FastTree using the Jukes-Cantor model of nucleotide evolution. Well-supported branches are illustrated. The* H. erato *gene* Herato0606.245 *(GGPPS, shows high similarity to the GGPPS of the moth* Choristoneura fumiferana*) was used to root the tree.*


FARM
SARM

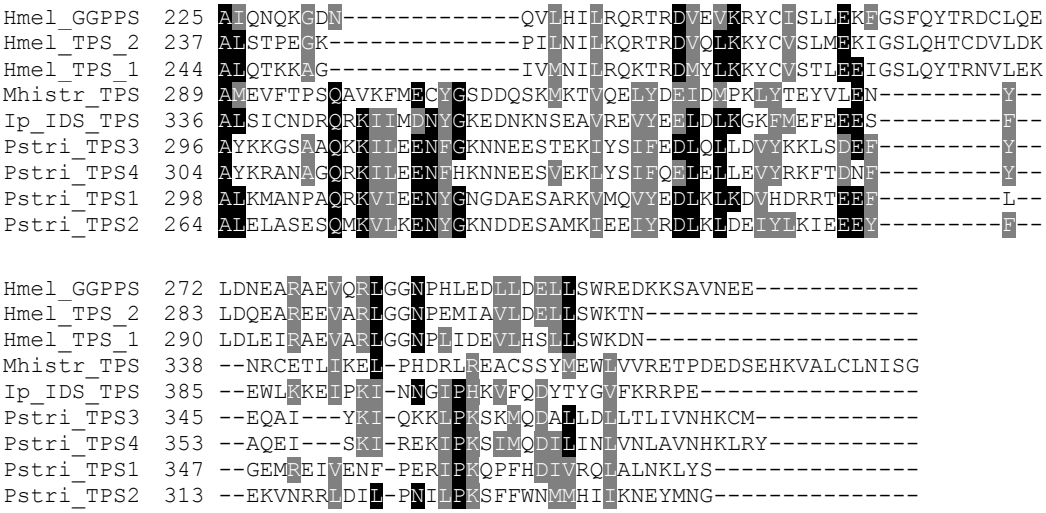


Figure S11: *Amino acid alignment between* H. melpomene *GGPPS and TPSs, and other insect TPSs. The two aspartate-rich motifs are labelled FARM and SARM. Stars show residues identified as conserved between insect TPSs (Lancaster et al. 2018). Hmel,* H. melpomene*; Mhistr,* Murgantia histrionica*; Ip*, Ips pini*; Pstri,* Phyllotreta striolata*; Hmel_TPS_1, HMEL037106g1; Hmel_TPS_2, HMEL037108g1.*

Table S1: Drosophila melanogaster *query protein sequences downloaded from FlyBase and searched (blastp) against all annotated proteins in the genome of* Heliconius melpomene *(v2.5) on LepBase to identify homologs of enzymes involved in the mevalonate and putative terpene synthesis pathways. The candidate orthologs identified in* H. melpomene *were then searched (blastp) against annotated proteins in the* D. melanogaster *genome on FlyBase. Reciprocal best blasts are highlighted in bold. We included other hits with an e-value smaller than* $1e^{-80}$.

Gene (symbol)	*D. melanogaster*	*H. melpomene*
Acetoacetyl-CoA thiolase (ACAT2)	CG9149	HMEL032609g1 HMEL014614g2 HMEL017484g1
Hydroxymethylglutaryl-CoA synthase (HMGCS)	CG4311	**HMEL005451g1**
Hydroxymethylglutaryl-CoA reductase (HMGCR)	CG10367	**HMEL016133g1**
Mevalonate kinase (MVK)	CG33671	**HMEL013262g2**
Phosphomevalonate kinase (PMVK)	CG10268	**HMEL007429g2/3***
Diphosphomevalonate decarboxylase (MVD)	CG8239	**HMEL004012g1**
Isopentenyl-diphosphate isomerase (IDI)	CG5919	**HMEL005103g1**
Farnesyl pyrophosphate synthase (FPPS)	CG12389	**HMEL017961g1** HMEL017961g2
Geranylgeranyl pyrophosphate synthase (GGPPS)	CG8593	**HMEL015484g1** HMEL037105g1 HMEL037106g1 HMEL022306g3 HMEL037107g1 HMEL037108g1
Decaprenyl pyrophosphate synthase subunit 1 (PDSS1)	CG31005	**HMEL016759g1**
Decaprenyl pyrophosphate synthase subunit 2 (PDSS2)	CG10585	**HMEL031784g1** HMEL011234g1 HMEL008172g1 HMEL008173g1

*The first two exons of *HMEL007429g2* and the last exon of *HMEL007429g3* are expressed as a single transcript:
(ATGGCACCAAAGATTGTATTATTGTTCAGTGGGAAAAGAAAATGTGGTAAAGACTTCGTGACTG
ATCATCTTAAGACATTGTTAAGTGACCAATGTGAAATTATAAAAATTTCACAACCCATCAAAAGTC

ATTGGGCAAAGGAAAAGAAATTAAATTTAAATGATCTCTTAAGTGATGGTGAATATAAAGAGAA
CTACCGCCTAGAAATGATAAAATGGAGTGAGGAAATGAGACAAAAAGATTATGGTTGTTTTTGTA
GAGCTGCATGTGAAAATGCTACAGAGAAACCTGTATGGATTGTCAGTGATATAAGACGGAAAAC
AGATTTGCAGTGGTTTAAAGAAACCTATGGTGATCTTATTAAAACAATTCGACTAACAGCAGATG
ACAATACTAGGACTGAAAGAGGTTTCCAATTTAAGAGTGGAGTTGATAATGCAGCCTCAGAATGT
GATTTAGACGATTATACAGAATGGGATCTCATTATTGAAAACAGCAAAGATAAAACTGTTGAGGA
TCTTACTAAAAATATTATACTTCTATTAGAATCTTTGAATATTTTACACAAACTTAGGTAG)

Table S2: *Transcript sequences for* H. cydno *genes from QTL region, as well as names used in* H. cydno *annotation. (Available at OSF: https://osf.io/3z9tg/?view_only=63ba7c0767a84d8eb907fbf599df062f).*

Table S3: *Linear model statistics for differential gene expression analysis in* H. melpomene *heads and abdomens of both sexes. The model includes two fixed terms, tissue and sex, their interaction, and a random term, individual (expression ~ sex + tissue + sex*tissue + (1|individual). The Log FC column gives the log2 Fold Change between the groups being compared, while the Ave. Expr. column gives the mean log2-exprtession across all samples. Column t is the moderated t-statistic and B is the B-statistic, the log odds that the gene is differentially expressed. The Adj. p-value column gives p-values (bold are significant) corrected for multiple testing using the Benjamini and Hochberg's method to control the false discovery rate across all tested genes (17,902).*

Gene	Term	Log FC	Ave. Expr.	t	p-value	Adj. p-value	B
HMEL016759g1	sex*tissue	0.6677	4.3438	1.3668	0.1863	0.4143	-6.1253
	sex	-0.7177	4.3438	-1.5627	0.1333	0.3131	-5.9173
	tissue	2.0102	4.3438	6.1544	4.41E-06	**3.14E-05**	3.7935
HMEL037105g1	sex*tissue	2.9238	-0.6946	2.4561	0.0230	0.0962	-3.7249
	sex	-6.1677	-0.6946	-7.2360	4.27E-07	**5.36E-06**	6.5817
	tissue	-5.3853	-0.6946	-7.5482	2.24E-07	**2.42E-06**	7.1670
HMEL037106g1	sex*tissue	-4.8560	3.0924	-4.3466	0.0003	**0.0028**	0.2480
	sex	3.9182	3.0924	7.6856	1.69E-07	**2.26E-06**	7.1353
	tissue	-5.0244	3.0924	-6.6000	1.65E-06	**1.36E-05**	5.1304
HMEL022306g3	sex*tissue	-1.7761	-2.6439	-2.0475	0.0535	0.1795	-4.3872
	sex	1.0489	-2.6439	1.4821	0.1534	0.3461	-5.4490
	tissue	-0.8095	-2.6439	-1.3163	0.2024	0.2995	-5.7741
HMEL037107g1	sex*tissue	-1.5469	-1.2854	-1.4056	0.1747	0.3976	-5.4444
	sex	0.2268	-1.2854	0.2483	0.8063	0.8941	-6.5585
	tissue	0.0553	-1.2854	0.0731	0.9424	0.9609	-6.7482
HMEL037108g1	sex*tissue	-1.9477	1.1227	-1.7246	0.0995	0.2750	-5.2255
	sex	2.0789	1.1227	2.3922	0.0263	0.0953	-4.2979
	tissue	-0.2850	1.1227	-0.3439	0.7344	0.8065	-6.8774
HMEL015484g1	sex*tissue	0.4644	4.0455	1.1758	0.2530	0.5043	-6.3477
	sex	-0.2477	4.0455	-0.6846	0.5012	0.6873	-6.8918
	tissue	1.3734	4.0455	5.0788	0.0001	**0.0003**	1.3380

Table S4: *Linear model statistics for differential gene expression analysis in* H. melpomene *and* H. cydno *abdomens of both sexes. The model includes two fixed terms, species and sex, and their interaction (expression ~ sex + species + species*tissue). The Log FC column gives the log2 Fold Change between the groups being compared, while the Ave. Expr. column gives the mean log2-exprtession across all samples. Column t is the moderated t-statistic and B is the B-statistic, the log odds that the gene is differentially expressed. The Adj. p-value column gives p-values (bold are significant) corrected for multiple testing using the Benjamini and Hochberg's method to control the false discovery rate across all tested genes (11,571).*

Gene	Term	Log FC	Ave. Expr.	t	p-value	Adj. p-value	B
HMEL016759g1	species*sex	-0.3223	5.4126	-1.4740	0.1558	0.4349	-6.6939
	species	1.7233	5.4126	7.8810	1.32E-07	**4.59E-07**	6.6022
	sex	0.0672	5.4126	0.3071	0.7619	0.8445	-8.1198
HMEL037105g1	species*sex	-0.1478	2.5652	-0.7547	0.4591	0.7264	-7.0952
	species	0.6662	2.5652	3.4010	0.0028	**0.0061**	-3.1192
	sex	2.9925	2.5652	15.2781	1.33E-12	**3.14E-11**	18.9110
HMEL037106g1	species*sex	0.8480	6.7288	3.1511	0.0050	**0.0446**	-3.7029
	species	-0.7377	6.7288	-2.7414	0.0125	**0.0236**	-5.1106
	sex	-1.0794	6.7288	-4.0113	0.0007	**0.0027**	-2.1200
HMEL022306g3	species*sex	-0.6182	0.9778	-1.3147	0.2033	0.4950	-6.4067
	species	2.2782	0.9778	4.8446	0.0001	**0.0003**	0.3653
	sex	-1.1116	0.9778	-2.3639	0.0282	0.0699	-5.0829
HMEL037107g1	species*sex	-0.2153	1.4275	-0.7773	0.4459	0.7173	-7.0103
	species	1.8967	1.4275	6.8481	1.09E-06	**3.53E-06**	4.8657
	sex	-0.3254	1.4275	-1.1747	0.2537	0.3944	-7.0160
HMEL037108g1	species*sex	1.1927	1.5270	3.4427	0.0025	**0.0257**	-2.4568
	species	-0.8135	1.5270	-2.3483	0.0291	**0.0503**	-5.2982
	sex	0.1373	1.5270	0.3963	0.6960	0.7963	-7.6400
HMEL015484g1	species*sex	-0.2868	4.0561	-3.8325	0.0010	**0.0123**	-1.9266
	species	0.2741	4.0561	3.6640	0.0015	**0.0035**	-2.8338
	sex	-0.1316	4.0561	-1.7586	0.0937	0.1854	-6.5949

Table S5: *Primer sequences. CACC was added to the 5' end of the forward primer so that it was compatible with the plasmid vector.*

Gene	Primer sequence	Use
HMEL037106g1	Forward: CACCATGTCAGAAACAGAAGTCC Reverse: TTAATTATCCTTCCAACTTAAAAGCGA	Amplification of transcript from cDNA library
HMEL037108g1	Forward: CACCATGGACGTTCAGAAAATAAGC Reverse: TTAATTCGTTTTCCAAGAAAGAAGTTC	Amplification of transcript from cDNA library
HMEL037106g1	Forward: GTTAATTCGTTACACGTAGC Reverse: TAATCTGGAAATAGCGACC	Sequencing PCR products
HMEL037108g1	Forward: CTAACTCTTAACGCCTCG Reverse: TTATATCCTCGCAGAAATCG	Sequencing PCR products
Both	Forward: AATACGACTCACTATAGGGG Reverse: GGTTAGGGATAGGCTTACC	Sequencing insert in plasmid

Table S6: *Quantification of experiments characterising TPS activity of HMEL037106g1 (Fig. 5A, Fig. S4). HMEL037106g1 is a monoterpene synthase, catalysing the formation of (*E*)-β-ocimene from GPP. Residual IDS activity is shown by the production of (*E*)-β-ocimene, linalool, and nerolidol from DMAPP and IPP. Mean amounts (ng) ± standard deviation for each compound across 3 replicates are shown. (control) indicates experiments where protein expression was not induced. N=3 for each treatment.*

	(*E*)-β-Ocimene	(*Z*)-β-Ocimene	Linalool	Geraniol	Nerolidol
DMAPP + IPP	7.8±1.4	0±0	3.4±0.4	0±0	4.4±1.8
DMAPP + IPP (control)	0±0	0±0	0±0	0±0	0±0
GPP + IPP	334.7±32.7	10.7±1.5	84.2±7.7	36±5.2	0±0
GPP + IPP (control)	0±0	0±0	17.2±2.6	17.4±4.6	0±0
GPP	356.5±115.3	12.0±4.4	108.4±37.3	66.6±23	0±0
GPP (control)	0±0	0±0	10.4±9	18.5±6	0±0
FPP + IPP	0±0	0±0	0±0	0±0	12.3±0.8
FPP + IPP (control)	0±0	0±0	0±0	0±0	2.1±0.3

Table S7: *Quantification of experiments characterising TPS activity of HMEL037108g1 (Fig. 5A, Fig. S4). HMEL037108g1 acts as a mono- and sesquiterpene synthase, producing linalool from GPP and nerolidol from FPP. Small amounts of linalool and nerolidol detected in DMAPP and IPP treatment, and of nerolidol in the GPP treatment, demonstrate residual IDS activity. Mean amounts (ng) ± standard deviation for each compound across 3 replicates are shown. N=3 for each treatment.*

	(*E*)-β-Ocimene	Linalool	Geraniol	Nerolidol	Farnesol
DMAPP + IPP	0±0	8.5±0.5	0±0	10.2±0.5	0±0
DMAPP + IPP (control)	0±0	0±0	0±0	0±0	0±0
GPP + IPP	11.0±1.4	2908.3±361.4	109.3±21.6	0±0	0±0
GPP + IPP (control)	0±0	44.5±3.0	63.6±1.7	0±0	0±0
GPP	16.8±1.5	4040.0±404.1	122.2±15.8	11.1±1.0	0±0
GPP (control)	0±0	40.8±5.1	57.8±1.2	0±0	0±0
FPP + IPP	0±0	0±0	0±0	1734.9±165.7	21.5±3.3
FPP + IPP (control)	0±0	0±0	0±0	4.0±0.2	0±0

Table S8: *HMEL037106g1 does not show enzymatic activity with linalool, demonstrating it is not an intermediate in the synthesis of (*E*)-β-ocimene (Fig. S5). Mean amounts (ng) ± standard deviation for each compound across 3 replicates are shown. N=3 for each treatment.*

	(*E*)-β-Ocimene	(*Z*)-β-Ocimene	Linalool
(*S*)-Linalool	0±0	0±0	3182.4±445.5
(*S*)-Linalool (control)	0±0	0±0	2698.9±1020.8
(*R*)-Linalool	0±0	0±0	3226.6±713
(*R*)-Linalool (control)	0±0	0±0	3275.8±350.3

Table S9: *Summary of products from enzymatic assays using precursors from different steps in the pathway (Fig. 1) with both HMEL037106g1 and HMEL037108g1 (Fig. 5A, Table S6, Table S7).*

Precursors	Enzyme	Products	Activity type
DMAPP + IPP	HMEL037106g1	Trace (*E*)-β-Ocimene Trace linalool Trace nerolidol	Residual GPS Monoterpene synthase Sesquiterpene synthase
DMAPP + IPP	HMEL037108g1	Trace linalool Trace nerolidol	Residual GPS Monoterpene synthase Sesquiterpene synthase
GPP + IPP	HMEL037106g1	(*E*)-β-Ocimene Trace (Z)-β-Ocimene Linalool	Monoterpene synthase
GPP + IPP	HMEL037108g1	Trace (*E*)-β-Ocimene Linalool	Monoterpene synthase
GPP	HMEL037106g1	(*E*)-β-Ocimene Trace (Z)-β-Ocimene Linalool	Monoterpene synthase
GPP	HMEL037108g1	Trace (*E*)-β-Ocimene Linalool	Monoterpene synthase
FPP + IPP	HMEL037106g1	None	None
FPP + IPP	HMEL037108g1	Nerolidol	Sesquiterpene synthase

Table S10: *Quantification of experiments characterising IDS activity of HMEL037106g1 (Fig. S7). Only residual IDS activity is detected, with small amounts of (*E*)-ß-ocimene, linalool, and nerolidol produced from DMAPP and IPP. No other IDS activity is detected. High amounts of geraniol and farnesol in both experimental and control treatments is due to dephosphorylation of GPP and FPP, respectively. The main function of HMEL037106g1 is the production of (*E*)-ß-ocimene from GPP. Mean amounts (ng) ± standard deviation for each compound across 3 replicates are shown. N=3 for each treatment.*

	(*E*)-β-Ocimene	(*Z*)-β-Ocimene	Linalool	Geraniol	Nerolidol	Farnesol
DMAPP + IPP	1.7±0.5	0±0	1±0.1	0±0	1.6±0.1	0±0
DMAPP + IPP (control)	0±0	0±0	0±0	0±0	0±0	0±0
GPP + IPP	325.3±17.4	11.3±0.7	109.2±10.5	1590.1±133.1	0±0	0±0
GPP + IPP (control)	2.9±0.2	0±0	17.6±0.8	2300.4±156	0±0	0±0
FPP + IPP	0±0	0±0	0±0	0±0	15.7±5.1	1320.3±114.5
FPP + IPP (control)	0±0	0±0	0±0	0±0	4±0.2	1582.0±65.6

Table S11: *Quantification of experiments characterising IDS activity of HMEL037108g1 (Fig. S8). TPS activity is again demonstrated by the production of linalool from FPP, and nerolidol from FPP. Only residual IDS activity is detected, by the presence of linalool and nerolidol in treatments with DMAPP and IPP, and nerolidol in the GPP treatment. Geraniol and farnesol are present due to dephosphorylation of remaining GPP and FPP in treatments. Mean amounts (ng) ± standard deviation for each compound across 3 replicates are shown. N=3 for each treatment.*

	(*E*)-β-Ocimene	Linalool	Geraniol	Nerolidol	Farnesol
DMAPP + IPP	0±0	3.9±1.6	0±0	3.5±0.9	0±0
DMAPP + IPP (control)	0±0	0±0	0±0	0±0	0±0
GPP + IPP	12.0±0.8	3208.4±261.5	290.0±24.5	3.6±0.7	0±0
GPP + IPP (control)	0±0	30.6±0.7	2117.0±184.9	0±0	0±0
FPP + IPP	0±0	0±0	0±0	1536.3±61.3	983.8±57.5
FPP + IPP (control)	0±0	0±0	0±0	4.9±5.0	1220.8±1105.8

Table S12: *Full names of species from Figure 6 in the main text.*

Abbreviation	Full name
A. gossypii	*Aphis gossypii*
A. grandis	*Anthonomus grandis*
A. thaliana	*Arabidopsis thaliana*
B. mori	*Bombyx mori*
B. terrestris	*Bombus terrestris*
C. fumiferana	*Choristoneura fumiferana*
C. reinhardtii	*Chlamydomonas reinhardtii*
C. unshiu	*Citrus reinhardtii*
D. melanogaster	*Drosophila melanogaster*
D. ponderosae	*Dendroctonus ponderosae*
F. fujikuroi	*Fusarium fujikuroi*
G. arboreum	*Gossypium arboreum*
G. biloba	*Ginkgo biloba*
H. lupulus	*Humulus lupulus*
H. melpomene	*Heliconius melpomene*
H. sapiens	*Homo sapiens*
I. pini	*Ips pini*
M. chamomilla	*Matricaria chamomilla*
M. domestica	*Malus domestica*
M. histrionica	*Murgantia histrionica*
M. lewisii	*Mimulus lewisii*
M. persicae	*Myzus persicae*
M. piperita	*Mentha piperita*
N. tabacum	*Nicotiana tabacum*
N. viridula	*Nezara viridula*
P. cochleariae	*Phaedon cochleariae*
P. striolata	*Phyllotreta striolata*
R. speratus	*Reticulitermes speratus*
S. cerevisiae	*Saccharomyces cerevisiae*
T. castaneum	*Tribolium castaneum*

Conclusion

Over the past 150 years we have learnt a great deal about adaptation and speciation from *Heliconius* butterflies (Merrill et al. 2015). One of the most impressive features of these butterflies is the wide diversity of wing colour patterns, even within a single species and research has focused on the evolution and more recently genetics of these patterns. In spite of a human bias for visual signals, the likely importance of chemical signalling in *Heliconius* has long been noted, but is much less well understood (Müller 1912a,b; Eltringham 1925; Barth and Barth 1952; Crane 1955; Emsley 1963). With a wealth of ecological and genetic studies, *Heliconius* is an excellent system to investigate the evolution of chemical signalling and its role in mate choice.

Behavioural observations suggested that male androconial compounds are important during courtship, however experimental data was previously lacking (Crane 1955; Klein and Araújo 2010). Female choice experiments, presented here in Chapter 2, demonstrate that chemical signalling is important for female choice and mating in *Heliconius*. Female *H. melpomene*, *H. erato* and *H. timareta* discriminate against males which have their androconia experimentally blocked, confirming the importance of male androconial regions for courtship in *Heliconius*. Morphological analyses also presented in this chapter describe the androconia, showing the brush-like scales found in this region. Chemical analyses then identify putative male sex pheromones of *H. melpomene*. These compounds are found in androconial regions of mature male *H. melpomene*, but absent in immature males and females.

Having established that androconial compounds are important for courtship and mating, the next question concerns the information conveyed by these chemical signals. I hypothesised that if chemical signalling is important for mate assessment, it could be a condition-dependent trait. One way to manipulate the condition of an individual is through dietary changes. Larval diet could affect which compounds, or pre-cursors, are sequestered, which might affect pheromone production. Both larval diet, and access to pollen as an adult, could affect individual condition. In Chapter 3 I present the results of manipulation of both the larval and adult diet of *H. melpomene*. Access to pollen does not

affect pheromone production, whilst larval diet affects minor, but not major, components of the chemical profile. These results suggest that whilst host plant affects the chemical profile of males, there must be other factors involved in determining intraspecific variation.

What remains unclear is whether changing minor components of the male chemical profile, as seen when changing larval diet, affects female choice. The most abundant compounds are not always the most important, as minor components can be important for attraction (D'Alessandro et al. 2009; McCormick et al. 2014) . The compounds which are significantly different between treatments could, therefore, affect female choice. This highlights the gap in knowledge which remains regarding the biological activity of the components of the bouquet. Whilst ideally we would carry out behavioural assays, this is not feasible for large numbers of species and populations. In Chapter 4, I propose that we can use statistical methods to try and identify male pheromone candidates. To do this, I studied the chemical profiles of the androconia and genitals of seven *Heliconius* species across a broad geographic range. I identify species-specific compounds, found in all populations of a given species, which I propose are potential male pheromones. This approach has proven promising as the compounds identified for *H. melpomene* androconia and genitals, octadecanal and (*E*)-β-ocimene, respectively, are known to be biologically active (Schulz et al. 2008; Byers et al. 2019). This statistical approach can therefore provide us with a starting point for more targeted behavioural trials using the candidates.

As well as trying to identify species-specific compounds, Chapter 4 also attempts to describe intra- and interspecific divergence to understand the evolution of chemical profiles in *Heliconius*. I find similar patterns to those previously seen for orchid bees (Weber et al. 2016). Species identity is the best predictor of chemical divergence, with larger inter- than intraspecific differences in chemical profiles. These species differences are consistent across a large geographic range, suggesting they could be involved in reproductive isolation. In the same chapter, I also investigated correlations between chemical divergence and genetic and geographic distance. In general, I find evidence that patterns of genetic divergence predict chemical divergence, consistent with neutral processes playing a role in profile evolution. These patterns are likely to be complicated

by the variety of processes driving their evolution. As well as neutral processes such as genetic drift, adaptive processes such as stabilising selection could drive profiles towards a species stereotype, whilst sexual selection might promote diversity.

Studying phenotypic patterns provides insight into the macroevolution of chemical profiles. This approach, however, does not reveal the genetic basis of the traits. Understanding the genetic basis of pheromone production not only helps us to study pheromone evolution, but also more general concepts in evolutionary biology, such as convergence and gene family evolution. Chapter 5 presents a study of the genetic basis of (*E*)-β-ocimene production, an anti-aphrodisiac pheromone of *H. melpomene*, combining a QTL study, gene expression analysis and a candidate gene approach, followed by *in vitro* expression and enzymatic assays of the promising candidates. I find an expansion in a family of enzymes annotated as geranylgeranyl pyrophosphate synthases (GGPPSs) in the *H. melpomene* genome, with terpene synthase (TPS) activity confirmed for two of these enzymes. As well as evidence for gene duplication, I also find pseudogenes in this group, suggesting that evolution within this enzyme family is highly dynamic. It is likely that the dynamics of gene duplication have facilitated the evolution of terpene synthesis in *Heliconius*. Phylogenetic analysis shows that these enzymes are unrelated to previously described insect and plant TPSs, representing an independent origin of TPS activity.

Future directions

There has been widespread interest in the complexities of mate choice as an example of multimodal communication (Partan 2013; Mitoyen et al. 2019). As well as male attraction to colour patterns, we now know that female choice based on chemical cues is important for mating in *Heliconius* butterflies. The few examples of female choice in *Heliconius* have focused solely on visual cues (Chouteau et al. 2017; Southcott and Kronforst 2018), and in Chapter 2 I focus solely on chemical cues. In *Bicyclus anynana*, both visual and chemical signalling are important for female choice (Costanzo and Monteiro 2007). In *Heliconius*, we also need to combine experiments with visual and chemical cues to test the relative importance of these cues for female mating decisions. Even in systems where a single trait is greatly exaggerated in one sex, the use of multiple cues for mate choice has been described (Candolin 2003; Chapman et al. 2017). It is likely

we will find that multiple cues are also important for mate choice in *Heliconius*, including chemical, visual, and behavioural cues.

Both male and female choice can play a role in reproductive isolation between pheromone strains of *Ostrinia nubilalis* (Dopman et al. 2010). In *Heliconius*, the importance of both female and male mate choice based on colour pattern has been demonstrated (Jiggins et al. 2001, 2004; Jiggins 2008; Merrill et al. 2014; Sánchez et al. 2015; Southcott and Kronforst 2018). The role of chemical signalling, however, has mainly been tested in the context of intraspecific mate choice. Interspecific "perfuming" experiments with *H. timareta* and *H. melpomene* suggest that either androconial compounds, or genital compounds, or both, influence female choice, therefore playing a role in reproductive isolation between these two species (Mérot *et al.,* 2015). I attempted to carry out similar experiments with *H. cydno* and *H. melpomene*, unfortunately without success. We need to test this both with other species pairs, and separately for androconial and genital compounds, to understand more about the role chemical signaling could play in interspecific mate choice and thus reproductive isolation in *Heliconius*.

As well as trying to understand the role of chemical signalling in mate choice and reproductive isolation, we need to determine which components of the chemical bouquets discovered are biologically active. Important first steps taken in this thesis have provided a list of candidate pheromones for different species, but for most species these still need to be confirmed behaviourally. In Chapter 4, I identified octadecanal as the candidate androconial pheromone of *H. melpomene*. More recent analyses have found it is bioactive, eliciting both an electroanntenographic and behavioural response (Byers et al. 2019). Similar experiments using electroanntenography and bioassays will be needed to test the candidates in other species.

We also need to understand more about how females perceive the bouquets as a whole. Minor components of bouquets can be important for attraction (D'Alessandro et al. 2009; McCormick et al. 2014) . The results from the larval diet experiments in Chapter 3 show that many minor compounds in the adult bouquet are affected by their host plant as a larva, however, it is unclear if females would be able to detect these differences, and

whether they would affect female choice. Furthermore, the concentration and ratio of components can also be important (Wang et al. 2010b), an aspect that has not been explored in this system. Both GC-EAD and behavioural assays will help us to understand the ability of females to distinguish between scents and test their preferences.

It will also be interesting to try and disentangle what information is being conveyed by chemical cues. It is likely that multiple components are biologically active, potentially conveying different types of information to females. We know that minor components have the potential to convey information about larval diet, but other factors could also be important. For example, we could investigate the effects of inbreeding and age, factors which affect male sex pheromone composition in *Bicyclus anynana* (Nieberding et al. 2012; van Bergen et al. 2013). Further experiments regarding condition dependency would be useful, perhaps extending the adult diet experiments with and without access to pollen for a month or longer. It could also be interesting to look at the ability of males to recover their bouquet after some of it is inevitably released during mating. Other approaches to test condition-dependence could also be trialled, such as changing the temperature during development, a technique that has been used to demonstrate the condition-dependence of iridescence in *Heliconius erato* (Melanie Brien, pers. comm.). Understanding which variables result in the observed intraspecific variation in chemical profiles is an important step in understand which types of information females could gain from chemical cues.

As well as focusing on the effect of different treatments on male pheromone composition itself, we can study the downstream benefits gained by females and how this correlates with male pheromone composition. Pheromones can signal indirect benefits to females, by increasing survivorship or mating success of offspring through provision of "good genes" (Andersson 1986, 1994). This seems to be the cases in male tobacco moths, *Ephestia elutella*, as large males which produce higher amounts of male pheromone also sire larger offspring (Phelan and Baker 1986). Alternatively, pheromones could signal the provision of a resource. This has been well-studied in the arctiid moth, *Uthetheisa ornatrix,* where the amount of male pheromone produced is an honest signal of alkaloids which are transferred to females during mating to chemically protect the eggs (Dussourd et al. 1988, 1991; Eisner and Meinwald 1995; Iyengar et al. 2001). *Heliconius* males

transfer a nutrient-rich spermatophore to females during mating which is used in the eggs and also female somatic tissue (Boggs and Gilbert 1979; Boggs 1981, 1990). It could be that male pheromone composition reflects the size or quality of the spermatophore (Cardoso and Silva 2015). Further experiments are needed to test the potential benefits to females signaled by male pheromones in *Heliconius.*

Ultimately, to understand how pheromones evolve and what drives their evolution we need to understand both their ecology and role in current populations, as well as their genetic basis. The discovery of a novel family of terpene synthase enzymes in *H. melpomene*, presented here in Chapter 5, has implications for understanding pheromone evolution not only in *Heliconius* but more broadly across insects. Within *Heliconius*, an interesting follow-up will be to determine the enzymatic activity of the other members of the enzyme family not tested in this thesis. Furthermore, assaying homologs from other *Heliconius* species will allow us to understand how these enzymes have evolved across *Heliconius* resulting in the different chemical profiles in each species. Pheromone differences between moth species have been attributed to both regulatory (Sakai et al. 2009; Albre et al. 2012) and coding changes (Lassance et al. 2010, 2013), as well as being linked to gene duplication events (Roelofs et al. 2002; Wang et al. 2010a). A comparative study in *Heliconius* could add to the current literature in Lepidoptera to understand the kinds of genetic changes which affect pheromone evolution.

Site-directed mutagenesis is another approach to determine the amino acid changes underlying differences in enzyme activity. In some cases a single amino acid change (Buček et al. 2015), or multiple changes (Lassance et al. 2013) can change the substrate specificity of an enzyme. The independent evolution of terpene synthases multiple times across insects provides an opportunity to study whether convergence has also occurred at the level of amino acid changes. This could be studied both by comparative studies with new examples from other insect lineages, as well as site-directed mutagenesis.

Concluding remarks

Chemical ecology is by nature an interdisciplinary area of research, combining both chemistry and ecology to understand how organisms interact with each other. This field, however, is much more than chemistry and ecology, combining behavioural ecology, genetics and genomics, neuroscience, and molecular biology, to gain a more complete understanding of the evolution of chemical signalling. In this thesis I have laid the foundation for some of these aspects of chemical ecology in *Heliconius*, opening doors for future research in this exciting area.

References

Abdallah, I. I., and W. J. Quax. 2017. A glimpse into the biosynthesis of terpenoids. KnE Life Sci. 81–98.

Ache, B., and J. Young. 2005. Olfaction: Diverse Species, Conserved Principles. Neuron 48:417–30.

Adams, H. B. and R. H. 2015. evobiR: Comparative and Population Genetic Analyses.

Albre, J., M. A. Liénard, T. M. Sirey, S. Schmidt, L. K. Tooman, C. Carraher, D. R. Greenwood, C. Löfstedt, and R. D. Newcomb. 2012. Sex Pheromone Evolution Is Associated with Differential Regulation of the Same Desaturase Gene in Two Genera of Leafroller Moths. PLOS Genet. 8:e1002489.

Ali, M. F., and E. D. Morgan. 1990. Chemical communication in insect communities: a guide to insect pheromones with special emphasis on social insects. Biol. Rev. 65:227–247.

Amo, L., and F. Bonadonna. 2018. Editorial: The Importance of Olfaction in Intra- and Interspecific Communication. Front. Ecol. Evol. 6.

Andersson, J., A.-K. Borg-Karlson, and C. Wiklund. 2000. Sexual cooperation and conflict in butterflies: a male–transferred anti–aphrodisiac reduces harassment of recently mated females. Proc. R. Soc. Lond. B Biol. Sci. 267:1271–1275.

Andersson, J., A.-K. Borg-Karlson, and C. Wiklund. 2004. Sexual conflict and anti-aphrodisiac titre in a polyandrous butterfly: male ejaculate tailoring and absence of female control. Proc. R. Soc. B Biol. Sci. 271:1765–1770.

Andersson, M. 1986. Evolution of Condition-Dependent Sex Ornaments and Mating Preferences: Sexual Selection Based on Viability Differences. Evolution 40:804–816.

Andersson, M. 1994. Sexual Selection. Princeton University Press, Princeton.

Andersson, M., and L. W. Simmons. 2006. Sexual selection and mate choice. Trends Ecol. Evol. 21:296–302.

Andersson, S., and H. E. M. Dobson. 2003. Antennal responses to floral scents in the butterfly *Heliconius melpomene*. J. Chem. Ecol. 29:2319–2330.

Andersson, S., L. A. Nilsson, I. Groth, and G. Bergström. 2002. Floral scents in butterfly-pollinated plants: possible convergence in chemical composition. Bot. J. Linn. Soc. 140:129–153.

Ando, T., S. Inomata, and M. Yamamoto. 2004. Lepidopteran Sex Pheromones. Pp. 51–96 in S. Schulz, ed. The Chemistry of Pheromones and Other Semiochemicals I. Springer Berlin Heidelberg.

Arn, H., E. Städler, and S. Rauscher. 1975. The Electroantennographic Detector — a Selective and Sensitive Tool in the Gas Chromatographic Analysis of Insect Pheromones. Z. Für Naturforschung C 30:722–725.

Assis, R., and D. Bachtrog. 2013. Neofunctionalization of young duplicate genes in *Drosophila*. Proc. Natl. Acad. Sci. 110:17409–17414.

Ayasse, M., J. Stökl, and W. Francke. 2011. Chemical ecology and pollinator-driven speciation in sexually deceptive orchids. Phytochemistry 72:1667–1677.

Bacquet, P. M. B., O. Brattström, H.-L. Wang, C. E. Allen, C. Löfstedt, P. M. Brakefield, and C. M. Nieberding. 2015. Selection on male sex pheromone composition contributes to butterfly reproductive isolation. Proc. R. Soc. Lond. B Biol. Sci. 282:20142734.

Bacquet, P. M. B., M. A. de Jong, O. Brattström, H.-L. Wang, F. Molleman, S. Heuskin, G. Lognay, C. Löfstedt, P. M. Brakefield, A. Vanderpoorten, and C. M. Nieberding. 2016. Differentiation in putative male sex pheromone components across and within populations of the African butterfly *Bicyclus anynana* as a potential driver of reproductive isolation. Ecol. Evol. 6:6064–6084.

Baker, T. C. 1989. Origin of courtship and sex pheromones of the oriental fruit moth and a discussion of the role of phytochemicals in the evolution of lepidopteran male scents. Pp. 401–418 in C. H. Chou and G. R. Waller, eds. Phytochemical Ecology: Allelochemicals, Mycotoxins, and Insect Pheromones and Allomones. Taipei, Taiwan.

Baker, T. C. 2002. Mechanism for saltational shifts in pheromone communication systems. Proc. Natl. Acad. Sci. U. S. A. 99:13368–13370.

Barbar, A., M. Couture, S. E. Sen, C. Béliveau, A. Nisole, M. Bipfubusa, and M. Cusson. 2013. Cloning, expression and characterization of an insect geranylgeranyl diphosphate synthase from *Choristoneura fumiferana*. Insect Biochem. Mol. Biol. 43:947–958.

Barth, R., and R. Barth. 1952. Os órgãos odoriferos masculinos de alguns Heliconiinae do Brasil. Mem. Inst. Oswaldo Cruz 50:335–422.

Bates, H. W. 1862. Contributions to an insect fauna of the Amazon valley. Lepidoptera: Heliconidae. Trans. Linn. Soc. Lond. 23:495–566.

Becker, H. G., and R. Beckert. 1993. Organikum-Organisch-Chemisches Praktikum. 22nd ed. Wiley-VCH Weinheim.

Bellés, X., D. Martín, and M.-D. Piulachs. 2005. The mevalonate pathway and the synthesis of juvenile hormone in insects. Annu. Rev. Entomol. 50:181–199.

Beltrán, M., C. D. Jiggins, V. Bull, M. Linares, J. Mallet, W. O. McMillan, and E. Bermingham. 2002. Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. Mol. Biol. Evol. 19:2176–2190.

Benedict, L., and R. C. K. Bowie. 2009. Macrogeographical variation in the song of a widely distributed African warbler. Biol. Lett. 5:484–487.

Benson, W. W. 1978. Resource partitioning in passion vine butterflies. Evol. Int. J. Org. Evol. 32:493–518.

Benson, W. W., K. S. Brown, and L. E. Gilbert. 1975. Coevolution of plants and herbivores: passion flower butterflies. Evol. Int. J. Org. Evol. 29:659–680.

Beran, F., T. G. Köllner, J. Gershenzon, and D. Tholl. 2019. Chemical convergence between plants and insects: biosynthetic origins and functions of common secondary metabolites. New Phytol. 223:52–67.

Beran, F., P. Rahfeld, K. Luck, R. Nagel, H. Vogel, N. Wielsch, S. Irmisch, S. Ramasamy, J. Gershenzon, D. G. Heckel, and T. G. Köllner. 2016. Novel family of terpene synthases evolved from trans-isoprenyl diphosphate synthases in a flea beetle. Proc. Natl. Acad. Sci. 113:2922–2927.

Bergstrom, C. T., and L. A. Real. 2000. Towards a theory of mutual mate choice: Lessons from two-sided matching. Evol. Ecol. Res. 2:493–508.

Bethenod, M.-T., Y. Thomas, F. Rousset, B. Frérot, L. Pélozuelo, G. Genestier, and D. Bourguet. 2004. Genetic isolation between two sympatric host plant races of the European corn borer, *Ostrinia nubilalis* Hübner. II: assortative mating and host-plant preferences for oviposition. Heredity 94:264–270.

Birch, M. C., G. M. Poppy, and T. C. Baker. 1990. Scents and Eversible Scent Structures of Male Moths. Annu. Rev. Entomol. 35:25–54.

Bittinger, K. 2017. usedist: Distance Matrix Utilities. R package version 0.1.0. https://CRAN.R-project.org/package=usedist.

Boerjan, W., J. Ralph, and M. Baucher. 2003. Lignin biosynthesis. Annu. Rev. Plant Biol. 54:519–546.

Boggs, C. L. 1981. Selection pressures affecting male nutrient investment at mating in heliconiine butterflies. Evolution 35:931–940.

Boggs, C. L. 1990. A general model of the role of male-donated nutrients in female insects' reproduction. Am. Nat. 136:598–617.

Boggs, C. L., and L. E. Gilbert. 1979. Male contribution to egg production in butterflies: evidence for transfer of nutrients at mating. Science 206:83–4.

Bohlmann, J., G. Meyer-Gauen, and R. Croteau. 1998. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. Proc. Natl. Acad. Sci. 95:4126–4133.

Bolnick, D. I., and M. Kirkpatrick. 2012. The relationship between intraspecific assortative mating and reproductive isolation between divergent populations. Curr. Zool. 58:484–492.

Bonduriansky, R. 2001. The evolution of male mate choice in insects: a synthesis of ideas and evidence. Biol. Rev. 76:305–339.

Brand, P., I. A. Hinojosa-Díaz, R. Ayala, M. Daigle, C. L. Y. Obiols, T. Eltz, and S. R. Ramírez. 2019. An olfactory receptor gene underlies reproductive isolation in perfume-collecting orchid bees. bioRxiv 537423.

Briscoe, A. D., S. M. Bybee, G. D. Bernard, F. Yuan, M. P. Sison-Mangus, R. D. Reed, A. D. Warren, J. Llorente-Bousquets, and C.-C. Chiao. 2010. Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in *Heliconius* butterflies. Proc. Natl. Acad. Sci. 107:3628–3633.

Briscoe, A. D., A. Macias-Muñoz, K. M. Kozak, J. R. Walters, F. Yuan, G. A. Jamie, S. H. Martin, K. K. Dasmahapatra, L. C. Ferguson, J. Mallet, E. Jacquin-Joly, and C. D. Jiggins. 2013. Female Behaviour Drives Expression and Evolution of Gustatory Receptors in Butterflies. PLOS Genet. 9:e1003620.

Brooks, R., and D. J. Kemp. 2001. Can older males deliver the good genes? Trends Ecol. Evol. 16:308–313.

Brower, A. V. Z. 1996. A new mimetic species of *Heliconius* (Lepidoptera: Nymphalidae), from southeastern Colombia, revealed by cladistic analysis of mitochondrial DNA sequences. Zool. J. Linn. Soc. 116:317–332.

Brower, L. P., J. V. Z. Brower, and C. T. Collins. 1963. Experimental studies of mimicry. 7. Relative palatability and Mullerian mimicry among neotropical butterflies of the subfamily Heliconiinae. Zool. Sci. Contrib. N. Y. Zool. Soc. 48:65–84.

Brower, L. P., and M. A. Jones. 1965. Precourtship interaction of wing and abdominal sex glands in male *Danaus* butterflies. Proc. R. Entomol. Soc. Lond. Ser. Gen. Entomol. 40:147–151.

Brown, K. S. 1981. The Biology of *Heliconius* and Related Genera. Annu. Rev. Entomol. 26:427–456.

Buček, A., P. Matoušková, H. Vogel, P. Šebesta, U. Jahn, J. Weißflog, A. Svatoš, and I. Pichová. 2015. Evolution of moth sex pheromone composition by a single amino acid substitution in a fatty acid desaturase. Proc. Natl. Acad. Sci. 112:12586–12591.

Burdfield-Steel, E., H. Pakkanen, B. Rojas, J. A. Galarza, and J. Mappes. 2018. De novo Synthesis of Chemical Defenses in an Aposematic Moth. J. Insect Sci. 18:1–4.

Buser, H. R., H. Arn, P. Guerin, and S. Rauscher. 1983. Determination of double bond position in mono-unsaturated acetates by mass spectrometry of dimethyl disulfide adducts. Anal. Chem. 55:818–822.

Butenandt, A., R. Beckmann, D. Stamm, and E. Hecker. 1959. Über den Sexuallockstoff des Seidenspinners *Bombyx mori.* Reindarstellung und Konstitution. Z Naturforsch 14b:283–284.

Butlin, R. K., G. M. Hewitt, and S. F. Webb. 1985. Sexual selection for intermediate optimum in *Chorthippus brunneus* (Orthoptera: Acrididae). Anim. Behav. 33:1281–1292.

Bybee, S. M., F. Yuan, M. D. Ramstetter, J. Llorente-Bousquets, R. D. Reed, D. Osorio, and A. D. Briscoe. 2012. UV Photoreceptors and UV-Yellow Wing Pigments in *Heliconius* Butterflies Allow a Color Signal to Serve both Mimicry and Intraspecific Communication. Am. Nat. 179:38–51.

Byers, K. J. R. P., K. Darragh, J. Musgrove, D. A. Almeida, S. F. Garza, I. A. Warren, P. Rastas, M. Kučka, Y. F. Chan, R. M. Merrill, S. Schulz, W. O. McMillan, and C. D. Jiggins.

2019. A major locus controls a biologically active pheromone component in *Heliconius melpomene*. bioRxiv 739037.

Cáceres, M. D., and P. Legendre. 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90:3566–3574.

Campbell, P., B. Pasch, J. L. Pino, O. L. Crino, M. Phillips, and S. M. Phelps. 2010. Geographic variation in the songs of neotropical singing mice: testing the relative importance of drift and local adaptation. Evolution 64:1955–1972.

Candolin, U. 2003. The use of multiple cues in mate choice. Biol. Rev. 78:575–595.

Carde, R.T., and J. D. Allison. 2016. Variation in Moth Pheromones. Causes and Consequences. Pp. 25-39 in J. D. Allison and R. T. Carde, eds. Pheromone Communication in Moths: Evolution, Behavior, and Application. University of California Press, Berkeley.

Cardoso, M. Z., and L. E. Gilbert. 2013. Pollen feeding, resource allocation and the evolution of chemical defence in passion vine butterflies. J. Evol. Biol. 26:1254–1260.

Cardoso, M. Z., and E. S. Silva. 2015. Spermatophore quality and production in two *Heliconius* butterflies with contrasting mating systems. J. Insect Behav. 28:693–703.

Castro, É. C. P. de, M. Zagrobelny, M. Z. Cardoso, and S. Bak. 2018. The arms race between heliconiine butterflies and *Passiflora* plants – new insights on an ancient subject. Biol. Rev. 93:555–573.

Chai, P., and R. B. Srygley. 1990. Predation and the Flight, Morphology, and Temperature of Neotropical Rain-Forest Butterflies. Am. Nat. 135:748–765.

Challis, R. J., S. Kumar, K. K. K. Dasmahapatra, C. D. Jiggins, and M. Blaxter. 2016. Lepbase: the Lepidopteran genome database. bioRxiv 056994.

Chapman, N. C., P. Siriwat, J. Howie, A. Towlson, L. Bellamy, K. Fowler, and A. Pomiankowski. 2017. The complexity of mating decisions in stalk‐eyed flies. Ecol. Evol. 7:6659–6668.

Chemnitz, J., P. C. Jentschke, M. Ayasse, and S. Steiger. 2015. Beyond species recognition: somatic state affects long-distance sex pheromone communication. Proc R Soc B 282:20150832.

Chen, X., T. G. Köllner, Q. Jia, A. Norris, B. Santhanam, P. Rabe, J. S. Dickschat, G. Shaulsky, J. Gershenzon, and F. Chen. 2016. Terpene synthase genes in eukaryotes beyond plants and fungi: Occurrence in social amoebae. Proc. Natl. Acad. Sci. U. S. A. 113:12132–12137.

Chen, F., D. Tholl, J. Bohlmann, and E. Pichersky. 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J. Cell Mol. Biol. 66:212–229.

Cheng, D., M. Meng, J. Peng, W. Qian, L. Kang, and Q. Xia. 2014. Genome-wide comparison of genes involved in the biosynthesis, metabolism, and signaling of juvenile hormone between silkworm and other insects. Genet. Mol. Biol. 37:444–459.

Chouteau, M., V. Llaurens, F. Piron-Prunier, and M. Joron. 2017. Polymorphism at a mimicry supergene maintained by opposing frequency-dependent selection pressures. Proc. Natl. Acad. Sci. 114:8325–8329.

Christianson, D. W. 2006. Structural biology and chemistry of the terpenoid cyclases. Chem. Rev. 106:3412–3442.

Clark, D. C., S. J. DeBano, and A. J. Moore. 1997. The influence of environmental quality on sexual selection in *Nauphoeta cinerea* (Dictyoptera: Blaberidae). Behav. Ecol. 8:46–53.

Clearwater, J. R. 1975. Pheromone metabolism in male *Pseudaletia separata* (Walk.) and *Mamestra configurata* (Walk.) (Lepidoptera: Noctuidae). Comp. Biochem. Physiol. Part B Comp. Biochem. 50:77–82.

Clutton-Brock, T. H., and G. A. Parker. 1992. Potential Reproductive Rates and the Operation of Sexual Selection. Q. Rev. Biol. 67:437–456.

Clutton-Brock, T. H., and A. C. J. Vincent. 1991. Sexual selection and the potential reproductive rates of males and females. Nature 351:58.

Collenette, C. L. 1929. On the odour of two species of *Heliconius*. (In "Observations on the bionomics of the Lepidoptera of Matto Grosso, Brazil," by C.L. Collenette & G. Talbot). Trans. Entomol. Soc. Lond. 72:409–410.

Conant, G. C., and K. H. Wolfe. 2008. Turning a hobby into a job: how duplicated genes find new functions. Nat. Rev. Genet. 9:938–950.

Conner, W. E., T. Eisner, R. K. V. Meer, A. Guerrero, and J. Meinwald. 1981. Precopulatory sexual interaction in an arctiid moth *Utetheisa ornatrix*: role of a pheromone derived from dietary alkaloids. Behav. Ecol. Sociobiol. 9:227–235.

Conner, W. E., and V. K. Iyengar. 2016. Male pheromones in moths: Reproductive isolation, sexy sons, and good genes. Pp. 191–208 in J. D. Allison and R. T. Carde, eds. Pheromone Communication in Moths: Evolution, Behavior, and Application. University of California Press, Berkeley.

Conrad, T., R. J. Paxton, G. Assum, and M. Ayasse. 2018. Divergence in male sexual odor signal and genetics across populations of the red mason bee, *Osmia bicornis*, in Europe. PLOS ONE 13:e0193153.

Costanzo, K., and A. Monteiro. 2007. The use of chemical and visual cues in female choice in the butterfly *Bicyclus anynana*. Proc. R. Soc. B Biol. Sci. 274:845–851.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer, Sunderland, MA.

Crane, J. 1955. Imaginal behaviour of a Trinidad butterfly, *Heliconius erato hydara* Hewitson, with special reference to the social use of color. Zool. N. Y. 40:167–196.

Cunningham, E. J. A., and T. R. Birkhead. 1998. Sex roles and sexual selection. Anim. Behav. 56:1311–1321.

Czaczkes, T. J., C. Grüter, and F. L. W. Ratnieks. 2015. Trail Pheromones: An Integrative View of Their Role in Social Insect Colony Organization. Annu. Rev. Entomol. 60:581–599.

Czech, L., J. Huerta-Cepas, and A. Stamatakis. 2017. A critical review on the use of support values in tree viewers and bioinformatics toolkits. Mol. Biol. Evol. 34:1535–1542.

D'Alessandro, M., V. Brunner, G. von Mérey, and T. C. J. Turlings. 2009. Strong attraction of the parasitoid *Cotesia marginiventris* towards minor volatile compounds of maize. J. Chem. Ecol. 35:999–1008.

Danci, A., P. W. Schaefer, A. Schopf, and G. Gries. 2006. Species-specific close-range sexual communication systems prevent cross-attraction in three species of *Glyptapanteles* parasitic wasps (Hymenoptera: Braconidae). Biol. Control 39:225–231.

Darragh, K., K. J. R. P. Byers, R. M. Merrill, W. O. McMillan, S. Schulz, and C. D. Jiggins. 2019a. Male pheromone composition depends on larval but not adult diet in *Heliconius melpomene*. Ecol. Entomol. 44:397–405.

Darragh, K., G. Montejo-Kovacevich, K. M. Kozak, C. R. Morrison, C. M. E. Figueiredo, J. S. Ready, C. Salazar, M. Linares, K. J. R. P. Byers, R. M. Merrill, W. O. McMillan, S. Schulz, and C. D. Jiggins. 2019b. Species specificity and intraspecific variation in the chemical profiles of *Heliconius* butterflies across a large geographic range. bioRxiv 573469.

Darragh, K., S. Vanjari, F. Mann, M. F. Gonzalez-Rojas, C. R. Morrison, C. Salazar, C. Pardo-Diaz, R. M. Merrill, W. O. McMillan, S. Schulz, and C. D. Jiggins. 2017. Male sex pheromone components in *Heliconius* butterflies released by the androconia affect female choice. PeerJ 5:e3953.

Darwin, C. 1863. [Review of] Contributions to an insect fauna of the Amazon Valley. By Henry Walter Bates, Esq. Transact. Linnean Soc. Vol. XXIII. 1862, p. 495. Nat. Hist. Rev. 3:219–224.

Davey, J. W., S. L. Barker, P. M. Rastas, A. Pinharanda, S. H. Martin, R. Durbin, W. O. McMillan, R. M. Merrill, and C. D. Jiggins. 2017. No evidence for maintenance of a sympatric *Heliconius* species barrier by chromosomal inversions. Evol. Lett. 1:138–154.

Davey, J. W., M. Chouteau, S. L. Barker, L. Maroja, S. W. Baxter, F. Simpson, R. M. Merrill, M. Joron, J. Mallet, K. K. Dasmahapatra, and C. D. Jiggins. 2016. Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. G3: Genes, Genomes, Genetics 6:695–708.

Degenhardt, J., T. G. Köllner, and J. Gershenzon. 2009. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry 70:1621–1637.

Degner, J. F., J. C. Marioni, A. A. Pai, J. K. Pickrell, E. Nkadori, Y. Gilad, and J. K. Pritchard. 2009. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. Bioinformatics 25:3207–3212.

Deinert, E. I., J. T. Longino, and L. E. Gilbert. 1994. Mate competition in butterflies. Nature 370:23.

Dettner, K., and C. Liepert. 1994. Chemical Mimicry and Camouflage. Annu. Rev. Entomol. 39:129–154.

Dinno, A. 2017. dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. R package version 1.3.4. https://CRAN.R-project.org/package=dunn.test.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

Dopman, E. B., P. S. Robbins, and A. Seaman. 2010. Components of reproductive isolation between North American pheromone strains of the European corn borer. Evol. Int. J. Org. Evol. 64:881–902.

Dorai-Raj, S. 2014. binom: Binomial Confidence Intervals For Several Parameterizations. R package version 1.1-1. https://CRAN.R-project.org/package=binom.

Dray, S., and A. B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22:1–20.

Dunlap-Pianka, H., C. L. Boggs, and L. E. Gilbert. 1977. Ovarian dynamics in heliconiine butterflies: programmed senescence versus eternal youth. Science 197:487–490.

Dussourd, D. E., C. A. Harvis, J. Meinwald, and T. Eisner. 1991. Pheromonal advertisement of a nuptial gift by a male moth (*Utetheisa ornatrix*). Proc. Natl. Acad. Sci. 88:9224–9227.

Dussourd, D. E., K. Ubik, C. Harvis, J. Resch, J. Meinwald, and T. Eisner. 1988. Biparental defensive endowment of eggs with acquired plant alkaloid in the moth *Utetheisa ornatrix*. Proc. Natl. Acad. Sci. U. S. A. 85:5992–5996.

Edde, P. A., T. W. Phillips, J. B. Robertson, and J. W. Dillwith. 2007. Pheromone Output by *Rhyzopertha dominica* (Coleoptera: Bostrichidae), as Affected by Host Plant and Beetle Size. Ann. Entomol. Soc. Am. 100:83–90.

Edwards, W. H. 1881. On certain habits of *Heliconia charitonia* Linn., a species of butterfly found in Florida. Papilio 1:209–215.

Eggert, A.-K., and J. K. Müller. 1997. Biparental care and social evolution in burying beetles: lessons from the larder. Pp. 216–236 in J. C. Choe and B. J. Crespi, eds. The Evolution of Social Behavior in Insects and Arachnids. Cambridge University Press, Cambridge.

Eirín-López, J. M., L. Rebordinos, A. P. Rooney, and J. Rozas. 2012. The birth-and-death evolution of multigene families revisited. Repetitive DNA 7:170–196.

Eisenreich, W., A. Bacher, D. Arigoni, and F. Rohdich. 2004. Biosynthesis of isoprenoids via the non-mevalonate pathway. Cell. Mol. Life Sci. CMLS 61:1401–1426.

Eisner, T., and J. Meinwald. 1995. The chemistry of sexual selection. Proc. Natl. Acad. Sci. 92:50–55.

Eisner, T., and J. Meinwald. 2003. Alkaloid-derived pheromones and sexual selection in Lepidoptera. Pp. 341–368 in G. J. Blomquist and R. G. Vogt, eds. Insect Pheromone Biochemistry and Molecular Biology. Elsevier/Academic Press, Amsterdam; Boston, MA.

El-Sayed, A. 2007. The Pherobase: Database of Insect Pheromones and Semiochemicals.

Eltringham, H. 1925. On the Abdominal Glands in *Heliconius* (Lepidoptera). Trans. R. Entomol. Soc. Lond. 73:269–275.

Eltz, T., W. M. Whitten, D. W. Roubik, and K. E. Linsenmair. 1999. Fragrance Collection, Storage, and Accumulation by Individual Male Orchid Bees. J Chem Ecol 25:157–176.

Emlen, S. T., and L. W. Oring. 1977. Ecology, sexual selection, and the evolution of mating systems. Science 197:215–223.

Emsley, M. 1963. A morphological study of imagine Heliconiinae (Lep.: Nymphalidae) with a consideration of the evolutionary relationships within the group. Zool. N. Y. 48:85–130.

Engler-Chaouat, H. S., and L. E. Gilbert. 2006. De novo synthesis vs. sequestration: negatively correlated metabolic traits and the evolution of host plant specialization in cyanogenic butterflies. J. Chem. Ecol. 33:25–42.

Estrada, C., and L. E. Gilbert. 2010. Host plants and immatures as mate-searching cues in *Heliconius* butterflies. Anim. Behav. 80:231–239.

Estrada, C., and C. D. Jiggins. 2008. Interspecific sexual attraction because of convergence in warning colouration: is there a conflict between natural and sexual selection in mimetic species? J. Evol. Biol. 21:749–760.

Estrada, C., S. Schulz, S. Yildizhan, and L. E. Gilbert. 2011. Sexual selection drives the evolution of antiaphrodisiac pheromones in butterflies. Evol. Int. J. Org. Evol. 65:2843–2854.

Estrada, C., S. Yildizhan, S. Schulz, and L. E. Gilbert. 2010. Sex-specific chemical cues from immatures facilitate the evolution of mate guarding in *Heliconius* butterflies. Proc. R. Soc. B Biol. Sci. 277:407–413.

Farré-Armengol, G., I. Filella, J. Llusià, and J. Peñuelas. 2017. β-Ocimene, a Key Floral and Foliar Volatile Involved in Multiple Interactions between Plants and Other Organisms. Molecules 22:1148.

Ferreira, M., and J. W. H. Ferguson. 2002. Geographic variation in the calling song of the field cricket *Gryllus bimaculatus* (Orthoptera: Gryllidae) and its relevance to mate recognition and mate choice. J. Zool. 257:163–170.

Finkbeiner, S. D., A. D. Briscoe, and R. D. Reed. 2014. Warning signals are seductive: Relative contributions of color and pattern to predator avoidance and mate attraction in *Heliconius* butterflies. Evolution 68:3410–3420.

Finkbeiner, S. D., D. A. Fishman, D. Osorio, and A. D. Briscoe. 2017. Ultraviolet and yellow reflectance but not fluorescence is important for visual discrimination of conspecifics by *Heliconius erato*. J. Exp. Biol. 220:1267–1276.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.

Forsberg, J., and C. Wiklund. 1989. Mating in the afternoon: Time-saving in courtship and remating by females of a polyandrous butterfly *Pieris napi* L. Behav. Ecol. Sociobiol. 25:349–356.

Fox, J., and S. Weisberg. 2011. An {R} Companion to Applied Regression. Second. Sage, Thousand Oaks, CA.

Francke, W., and S. Schulz. 2010. Pheromones of Terrestrial Invertebrates. Pp. 153–223 in H.-W. Liu and L. Mander, eds. Comprehensive Natural Products II. Elsevier, Oxford.

Friberg, M., N. Vongvanich, A.-K. Borg-Karlson, D. J. Kemp, S. Merilaita, and C. Wiklund. 2008. Female mate choice determines reproductive isolation between sympatric butterflies. Behav. Ecol. Sociobiol. 62:873–886.

Gerhardt, H. C. 1982. Sound Pattern Recognition in Some North American Treefrogs (Anura: Hylidae): Implications for Mate Choice. Integr. Comp. Biol. 22:581–595.

Gershenzon, J., and N. Dudareva. 2007. The function of terpene natural products in the natural world. Nat. Chem. Biol. 3:408–414.

Gilbert, L. E. 1972. Pollen Feeding and Reproductive Biology of *Heliconius* Butterflies. Proc. Natl. Acad. Sci. 69:1403–1407.

Gilbert, L. E. 1976. Postmating female odor in *Heliconius* butterflies: a male-contributed antiaphrodisiac? Science 193:419–420.

Gilbert, L. E. 1991. Biodiversity of a Central American *Heliconius* community: pattern, process, and problems. Pp. 403–427 in P. W. Price, T. M. Lewinsohn, T. W. Fernandes, and W. W. Benson, eds. Plant-Animal Interactions: Evolutionary Ecology in Tropical and Temperate Regions. John Wiley, New York.

Gilg, A. B., J. C. Bearfield, C. Tittiger, W. H. Welch, and G. J. Blomquist. 2005. Isolation and functional expression of an animal geranyl diphosphate synthase and its role in bark beetle pheromone biosynthesis. Proc. Natl. Acad. Sci. 102:9760–9765.

Gilg, A. B., C. Tittiger, and G. J. Blomquist. 2009. Unique animal prenyltransferase with monoterpene synthase activity. Naturwissenschaften 96:731–735.

Giraldo, N., C. Salazar, C. D. Jiggins, E. Bermingham, and M. Linares. 2008. Two sisters in the same dress: *Heliconius* cryptic species. BMC Evol. Biol. 8:324.

Girard, M. B., D. O. Elias, and M. M. Kasumovic. 2015. Female preference for multi-modal courtship: multiple signals are important for male mating success in peacock spiders. Proc R Soc B 282:20152222.

Granero, A. M., J. M. G. Sanz, F. J. E. Gonzalez, J. L. M. Vidal, A. Dornhaus, J. Ghani, A. R. Serrano, and L. Chittka. 2005. Chemical compounds of the foraging recruitment pheromone in bumblebees. Naturwissenschaften 92:371–374.

Gray, D. A., and W. H. Cade. 2000. Sexual selection and speciation in field crickets. Proc. Natl. Acad. Sci. 97:14449–14454.

Grillet, M., L. Dartevelle, and J.-F. Ferveur. 2006. A *Drosophila* male pheromone affects female sexual receptivity. Proc. R. Soc. B Biol. Sci. 273:315–323.

Grillet, M., C. Everaerts, B. Houot, M. G. Ritchie, M. Cobb, and J.-F. Ferveur. 2012. Incipient speciation in *Drosophila melanogaster* involves chemical signals. Sci. Rep. 2:224.

Groot, A. T., T. Dekker, and D. G. Heckel. 2016. The Genetic Basis of Pheromone Evolution in Moths. Annu. Rev. Entomol. 61:99–117.

Grula, J. W., J. D. McChesney, and O. R. Taylor. 1980. Aphrodisiac pheromones of the sulfur butterflies *Colias eurytheme* and *C. Philodice* (Lepidoptera, Pieridae). J. Chem. Ecol. 6:241–256.

Guilford, T., C. Nicol, M. Rothschild, and B. P. Moore. 1987. The biological roles of pyrazines: evidence for a warning odour function. Biol. J. Linn. Soc. 31:113–128.

Hall, J. P. W., and D. J. Harvey. 2002. A survey of androconial organs in the Riodinidae (Lepidoptera). Zool. J. Linn. Soc. 136:171–197.

Hankison, S. J., and M. B. Ptacek. 2008. Geographical variation of genetic and phenotypic traits in the Mexican sailfin mollies, *Poecilia velifera* and *P. petenensis*. Mol. Ecol. 17:2219–2233.

Harari, A. R., T. Zahavi, and D. Thiéry. 2011. Fitness Cost of Pheromone Production in Signaling Female Moths. Evolution 65:1572–1582.

He, X. J., X. C. Zhang, W. J. Jiang, A. B. Barron, J. H. Zhang, and Z. J. Zeng. 2016. Starving honey bee (*Apis mellifera*) larvae signal pheromonally to worker bees. Sci. Rep. 6:22359.

Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98.

Henneken, J., J. Q. D. Goodger, T. M. Jones, and M. A. Elgar. 2017. Diet-Mediated Pheromones and Signature Mixtures Can Enforce Signal Reliability. Front. Ecol. Evol. 4:145.

Hernández-Roldán, J. L., R. Bofill, L. Dapporto, M. L. Munguira, and R. Vila. 2014. Morphological and chemical analysis of male scent organs in the butterfly genus *Pyrgus* (Lepidoptera: Hesperiidae). Org. Divers. Evol. 14:269–278.

Hervé, M. 2018. RVAideMemoire: Testing and Plotting Procedures for Biostatistics. R package version 0.9-69-3.

Hijmans, R. J. 2017. geosphere: Spherical Trigonometry. R package version 1.5-7.

Hillier, N. K., and N. J. Vickers. 2004. The role of heliothine hairpencil compounds in female *Heliothis virescens* (Lepidoptera: Noctuidae) behavior and mate acceptance. Chem. Senses 29:499–511.

Hillier, N. K., and N. J. Vickers. 2011. Hairpencil volatiles influence interspecific courtship and mating between two related moth species. J. Chem. Ecol. 37:1127–1136.

Huson, D. H., and C. Scornavacca. 2012. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. Syst. Biol. 61:1061–1067.

Irwin, D. E., M. P. Thimgan, and J. H. Irwin. 2008. Call divergence is correlated with geographic and genetic distance in greenish warblers (*Phylloscopus trochiloides*): a strong role for stochasticity in signal evolution? J. Evol. Biol. 21:435–448.

Itagaki, H., and W. E. Conner. 1988. Calling Behavior of *Manduca sexta* (L.) (Lepidoptera: Sphingidae) with Notes on the Morphology of the Female Sex Pheromone Gland. Ann. Entomol. Soc. Am. 81:798–807.

Iyengar, V. K., C. Rossini, and T. Eisner. 2001. Precopulatory assessment of male quality in an arctiid moth *Utetheisa ornatrix*: hydroxydanaidal is the only criterion of choice. Behav. Ecol. Sociobiol. 49:283–288.

Jiggins, C. D. 2008. Ecological Speciation in Mimetic Butterflies. BioScience 58:541–548.

Jiggins, C. D. 2017. The Ecology and Evolution of *Heliconius* Butterflies. Oxford University Press, Oxford.

Jiggins, C. D., C. Estrada, and A. Rodrigues. 2004. Mimicry and the evolution of premating isolation in *Heliconius melpomene* Linnaeus. J. Evol. Biol. 17:680–691.

Jiggins, C. D., R. E. Naisbit, R. L. Coe, and J. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411:302–305.

Johansson, B. G., and T. M. Jones. 2007. The role of chemical communication in mate choice. Biol. Rev. 82:265–289.

Johansson, B. G., T. M. Jones, and F. Widemo. 2005. Cost of pheromone production in a lekking *Drosophila.* Anim. Behav. 69:851–858.

Karino, K., T. Utagawa, and S. Shinjo. 2005. Heritability of the algal-foraging ability: an indirect benefit of female mate preference for males' carotenoid-based coloration in the guppy, *Poecilia reticulata*. Behav. Ecol. Sociobiol. 59:1.

Kassambara, A. 2019. ggpubr: "ggplot2" Based Publication Ready Plots. R package version 0.2.4. https://CRAN.R-project.org/package=ggpubr.

Kellogg, B. A., and C. D. Poulter. 1997. Chain elongation in the isoprenoid biosynthetic pathway. Curr. Opin. Chem. Biol. 1:570–578.

Kemp, D. J., and R. L. Rutowski. 2011. The Role of Coloration In Mate Choice And Sexual Interactions In Butterflies. Pp. 55–92 in Advances in the Study of Behavior. Elsevier, Amsterdam, Netherlands.

Kent, W. J. 2002. BLAT--the BLAST-like alignment tool. Genome Res. 12:656–664.

Klein, A. L., and A. M. de Araújo. 2010. Courtship behavior of *Heliconius erato phyllis* (Lepidoptera, Nymphalidae) towards virgin and mated females: conflict between attraction and repulsion signals? J. Ethol. 28:409–420.

Kock, D., J. Ruther, and K. P. Sauer. 2007. A Male Sex Pheromone in a Scorpionfly. J. Chem. Ecol. 33:1249–1256.

Kolmogorov, M., B. Raney, B. Paten, and S. Pham. 2014. Ragout—a reference-assisted assembly tool for bacterial genomes. Bioinformatics 30:i302–i309.

Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15:356.

Kozak, K. M., N. Wahlberg, A. F. E. Neild, K. K. Dasmahapatra, J. Mallet, and C. D. Jiggins. 2015. Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. Syst. Biol. 64:505–524.

Kronforst, M. R., L. G. Young, D. D. Kapan, C. McNeely, R. J. O'Neill, and L. E. Gilbert. 2006. Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. Proc. Natl. Acad. Sci. 103:6575–6580.

Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12.

Lai, C., R. McMahon, C. Young, T. F. C. Mackay, and C. H. Langley. 1998. quemao, a *Drosophila* bristle locus, encodes geranylgeranyl pyrophosphate Synthase. Genetics 149:1051–1061.

Lamas, G., and C. D. Jiggins. 2017. Taxonomic list. Pp. 214–244 in The ecology and evoution of *Heliconius* butterflies. Oxford University Press, Oxford.

Lancaster, J., A. Khrimian, S. Young, B. Lehner, K. Luck, A. Wallingford, S. K. B. Ghosh, P. Zerbe, A. Muchlinski, P. E. Marek, M. E. Sparks, J. G. Tokuhisa, C. Tittiger, T. G. Köllner, D. C. Weber, D. E. Gundersen-Rindal, T. P. Kuhar, and D. Tholl. 2018. De novo formation of an aggregation pheromone precursor by an isoprenyl diphosphate synthase-related terpene synthase in the harlequin bug. Proc. Natl. Acad. Sci. U. S. A. 115:E8634–E8641.

Lancaster, J., B. Lehner, A. Khrimian, A. Muchlinski, K. Luck, T. G. Köllner, D. C. Weber, D. E. Gundersen-Rindal, and D. Tholl. 2019. An IDS-type sesquiterpene synthase produces the pheromone precursor (*Z*)-α-Bisabolene in *Nezara viridula*. J. Chem. Ecol. 45:187–197.

Landolt, P. J., and T. W. Phillips. 1997. Host plant influences on sex pheromone behavior of phytophagous insects. Annu. Rev. Entomol. 42:371–391.

Larsdotter-Mellström, H., K. Eriksson, I. Liblikas I, C. Wiklund, A. K. Borg-Karlson, S. Nylin, N. Janz, and M. A. Carlsson. 2016. It's all in the mix: blend-specific behavioral response to a sexual pheromone in a butterfly. Front. Physiol. 7:68.

Lassance, J.-M., A. T. Groot, M. A. Liénard, B. Antony, C. Borgwardt, F. Andersson, E. Hedenström, D. G. Heckel, and C. Löfstedt. 2010. Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. Nature 466:486–489.

Lassance, J.-M., M. A. Liénard, B. Antony, S. Qian, T. Fujii, J. Tabata, Y. Ishikawa, and C. Löfstedt. 2013. Functional consequences of sequence variation in the pheromone biosynthetic gene pgFAR for *Ostrinia* moths. Proc. Natl. Acad. Sci. 110:3967–3972.

Lassance, J.-M., and C. Löfstedt. 2009. Concerted evolution of male and female display traits in the European corn borer, *Ostrinia nubilalis*. BMC Biol. 7:10.

Law, C. W., Y. Chen, W. Shi, and G. K. Smyth. 2014. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15:R29.

Leonhardt, S., C. Rasmussen, and T. Schmitt. 2013. Genes versus environment: Geography and phylogenetic relationships shape the chemical profiles of stingless bees on a global scale. Proc. Biol. Sci. 280:20130680.

Lespinet, O., Y. I. Wolf, E. V. Koonin, and L. Aravind. 2002. The Role of Lineage-Specific Gene Family Expansion in the Evolution of Eukaryotes. Genome Res. 12:1048–1059.

Lewis, S., S. Searle, N. Harris, M. Gibson, V. Iyer, J. Richter, C. Wiel, L. Bayraktaroglu, E. Birney, M. Crosby, J. Kaminker, B. Matthews, S. Prochnik, C. Smith, J. Tupy, G. Rubin, S. Misra, C. Mungall, and M. Clamp. 2002. Apollo: a sequence annotation editor. Genome Biology 3:research0082.1.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv13033997 Q-Bio.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. 25:1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. 25:2078–2079.

Liao, Y., G. K. Smyth, and W. Shi. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinforma. Oxf. Engl. 30:923–930.

Liedo, P., D. Orozco, L. Cruz-López, J. L. Quintero, C. Becerra-Pérez, M. del R. Hernández, A. Oropeza, and J. Toledo. 2013. Effect of post-teneral diets on the performance of sterile *Anastrepha ludens* and *Anastrepha obliqua* fruit flies. J. Appl. Entomol. 137:49–60.

Liénard, M. A., H.-L. Wang, J.-M. Lassance, and C. Löfstedt. 2014. Sex pheromone biosynthetic pathways are conserved between moths and the butterfly *Bicyclus anynana*. Nat. Commun. 5:3957.

Löfstedt, C. 1993. Moth Pheromone Genetics and Evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. 340:167–177.

Löfstedt, C., W. M. Herrebout, and S. B. Menken. 1991. Sex pheromones and their potential role in the evolution of reproductive isolation in small ermine moths (Yponomeutidae). Chemoecology 2:20–28.

Lundgren, L., and G. Bergström. 1975. Wing scents and scent-released phases in the courtship behavior of *Lycaeides argyrognomon* (Lepidoptera: Lycaenidae). J. Chem. Ecol. 1:399–412.

Madeira, F., Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn, and R. Lopez. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res., doi: 10.1093/nar/gkz268.

Magnard, J.-L., A. R. Bony, F. Bettini, A. Campanaro, B. Blerot, S. Baudino, and F. Jullien. 2018. Linalool and linalool nerolidol synthases in roses, several genes for little scent. Plant Physiol. Biochem. 127:74–87.

Maisonnasse, A., J.-C. Lenoir, D. Beslay, D. Crauser, and Y. L. Conte. 2010. E-β-Ocimene, a Volatile Brood Pheromone Involved in Social Regulation in the Honey Bee Colony (*Apis mellifera*). PLOS ONE 5:e13531.

Malinsky, M., J. T. Simpson, and R. Durbin. 2016. trio-sga: facilitating de novo assembly of highly heterozygous genomes with parent-child trios. bioRxiv 051516.

Mallet, J. 1986. Gregarious roosting and home range in *Heliconius* butterflies. Natl. Geogr. Res. 2:198–215.

Malouines, C. 2017. Counter-perfume: using pheromones to prevent female remating. Biol. Rev. 92:1570–1581.

Mann, F., S. Vanjari, N. Rosser, S. Mann, K. K. Dasmahapatra, C. Corbin, M. Linares, C. Pardo-Diaz, C. Salazar, C. Jiggins, and S. Schulz. 2017. The Scent Chemistry of *Heliconius* Wing Androconia. J. Chem. Ecol. 43:843–857.

Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, and S. H. Bryant. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43:D222-226.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10–12.

Martin, M. D., and T. C. Mendelson. 2016. The accumulation of reproductive isolation in early stages of divergence supports a role for sexual selection. J. Evol. Biol. 29:676–689.

Martin, S. H., M. Möst, W. J. Palmer, C. Salazar, W. O. McMillan, F. M. Jiggins, and C. D. Jiggins. 2016. Natural selection and genetic diversity in the butterfly *Heliconius melpomene*. Genetics 203:525–541.

Mas, F., and J.-M. Jallon. 2005. Sexual Isolation and Cuticular Hydrocarbon Differences between *Drosophila santomea* and *Drosophila yakuba*. J. Chem. Ecol. 31:2747–2752.

McCormick, A. C., J. Gershenzon, and S. B. Unsicker. 2014. Little peaks with big effects: establishing the role of minor plant volatiles in plant–insect interactions. Plant Cell Environ. 37:1836–1844.

McNeil, J. N., and J. Delisle. 1989. Are host plants important in pheromone-mediated mating systems of Lepidoptera? Experientia 45:236–240.

McPeek, M. A., L. B. Symes, D. M. Zong, and C. L. McPeek. 2011. Species recognition and patterns of population variation in the reproductive structures of a damselfly genus. Evol. Int. J. Org. Evol. 65:419–428.

Meinwald, J., Y. C. Meinwald, and P. H. Mazzocchi. 1969. Sex pheromone of the queen butterfly: chemistry. Science 164:1174–1175.

Melo, M. C., C. Salazar, C. D. Jiggins, and M. Linares. 2009. Assortative Mating Preferences Among Hybrids Offers a Route to Hybrid Speciation. Evolution 63:1660–1665.

Mendelson, T. C., and K. L. Shaw. 2012. The (mis)concept of species recognition. Trends Ecol. Evol. 27:421–427.

Menzel, F., R. Radke, and S. Foitzik. 2016. Odor diversity decreases with inbreeding in the ant *Hypoponera opacior*. Evolution 70:2573–2582.

Mérot, C., B. Frérot, E. Leppik, and M. Joron. 2015. Beyond magic traits: Multimodal mating cues in *Heliconius* butterflies. Evolution 69:2891–2904.

Mérot, C., J. Mavárez, A. Evin, K. K. Dasmahapatra, J. Mallet, G. Lamas, and M. Joron. 2013. Genetic differentiation without mimicry shift in a pair of hybridizing *Heliconius* species (Lepidoptera: Nymphalidae). Biol. J. Linn. Soc. 109:830–847.

Mérot, C., C. Salazar, R. M. Merrill, C. D. Jiggins, and M. Joron. 2017. What shapes the continuum of reproductive isolation? Lessons from *Heliconius* butterflies. Proc. Biol. Sci. 284.

Merrill, R. M., A. Chia, and N. J. Nadeau. 2014. Divergent warning patterns contribute to assortative mating between incipient *Heliconius* species. Ecol. Evol. 4:911–917.

Merrill, R. M., K. K. Dasmahapatra, J. W. Davey, D. D. Dell'Aglio, J. J. Hanly, B. Huber, C. D. Jiggins, M. Joron, K. M. Kozak, V. Llaurens, S. H. Martin, S. H. Montgomery, J. Morris, N. J. Nadeau, A. L. Pinharanda, N. Rosser, M. J. Thompson, S. Vanjari, R. W. R. Wallbank, and Q. Yu. 2015. The diversification of *Heliconius* butterflies: what have we learned in 150 years? J. Evol. Biol. 28:1417–1438.

Merrill, R. M., Z. Gompert, L. M. Dembeck, M. R. Kronforst, W. O. McMillan, and C. D. Jiggins. 2011a. Mate Preference Across the Speciation Continuum in a Clade of Mimetic Butterflies. Evolution 65:1489–1500.

Merrill, R. M., R. E. Naisbit, J. Mallet, and C. D. Jiggins. 2013. Ecological and genetic factors influencing the transition between host-use strategies in sympatric *Heliconius* butterflies. J. Evol. Biol. 26:1959–1967.

Merrill, R. M., P. Rastas, S. H. Martin, M. C. Melo, S. Barker, J. Davey, W. O. McMillan, and C. D. Jiggins. 2019. Genetic dissection of assortative mating behavior. PLOS Biol. 17:e2005902.

Merrill, R. M., B. Van Schooten, J. A. Scott, and C. D. Jiggins. 2011b. Pervasive genetic associations between traits causing reproductive isolation in *Heliconius* butterflies. Proc. R. Soc. B Biol. Sci. 278:511–518.

Mitoyen, C., C. Quigley, and L. Fusani. 2019. Evolution and function of multimodal courtship displays. Ethology 125:503–515.

Montgomery, S. H., R. M. Merrill, and S. R. Ott. 2016. Brain composition in *Heliconius* butterflies, posteclosion growth and experience-dependent neuropil plasticity. Journal of Comparative Neurology 524:1747–1769.

Moran, N. A., and T. Jarvik. 2010. Lateral Transfer of Genes from Fungi Underlies Carotenoid Production in Aphids. Science 328:624–627.

More, J. D., and N. S. Finney. 2002. A Simple and Advantageous Protocol for the Oxidation of Alcohols with o-Iodoxybenzoic Acid (IBX). Org. Lett. 4:3001–3003.

Mullen, L. M., S. N. Vignieri, J. A. Gore, and H. E. Hoekstra. 2009. Adaptive basis of geographic variation: genetic, phenotypic and environmental differences among beach mouse populations. Proc. R. Soc. Lond. B Biol. Sci. 276:3809–3818.

Müller, F. 1912. X. The Scent-scales of the Male “Maracujá butterflies.” Pp. 655–659 in G. B. Longstaff, ed. Butterfly hunting in many lands. Longmans, Green & Co., New York.

Müller, F. 1912a. IX. The “Maracujá (or Passion-flowers) butterflies”. Pp. 651–654 in G. B. Longstaff, ed. Butterfly hunting in many lands. Longmans, Green & Co., New York.

Muñoz, A. G., C. Salazar, J. Castaño, C. D. Jiggins, and M. Linares. 2010. Multiple sources of reproductive isolation in a bimodal butterfly hybrid zone. J. Evol. Biol. 23:1312–1320.

Nagegowda, D. A., M. Gutensohn, C. G. Wilkerson, and N. Dudareva. 2008. Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. Plant J. 55:224–239.

Nagel, L., and D. Schluter. 1998. Body Size, Natural Selection, and Speciation in Sticklebacks. Evolution 52:209–218.

Nahrstedt, A., and R. H. Davis. 1985. Biosynthesis and quantitative relationships of the cyanogenic glucosides, linamarin and lotaustralin, in genera of the Heliconiini (Insecta: Lepidoptera). Comp. Biochem. Physiol. Part B Comp. Biochem. 82:745–749.

Naisbit, R. E., C. D. Jiggins, and J. Mallet. 2001. Disruptive sexual selection against hybrids contributes to speciation between *Heliconius cydno* and *Heliconius melpomene.* Proc. R. Soc. Lond. Ser. B-Biol. Sci. 268:1849–1854.

Nei, M., X. Gu, and T. Sitnikova. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci. 94:7799–7806.

Nei, M., and A. P. Rooney. 2005. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39:121–152.

Nieberding, C. M., K. Fischer, M. Saastamoinen, C. E. Allen, E. A. Wallin, E. Hedenström, and P. M. Brakefield. 2012. Cracking the olfactory code of a butterfly: the scent of ageing. Ecol. Lett. 15:415–424.

Nieberding, C. M., H. de Vos, M. V. Schneider, J.-M. Lassance, N. Estramil, J. Andersson, J. Bång, E. Hedenström, C. Löfstedt, and P. M. Brakefield. 2008. The Male Sex Pheromone of the Butterfly *Bicyclus anynana*: Towards an Evolutionary Analysis. PLoS ONE 3:e2751.

Nishida, R., S. Schulz, C. S. Kim, H. Fukami, Y. Kuwahara, K. Honda, and N. Hayashi. 1996. Male sex pheromone of a giant danaine butterfly, *Idea leuconoe*. J. Chem. Ecol. 22:949–972.

Noriega, F. G., J. M. C. Ribeiro, J. F. Koener, J. G. Valenzuela, S. Hernandez-Martinez, V. M. Pham, and R. Feyereisen. 2006. Comparative genomics of insect juvenile hormone biosynthesis. Insect Biochem. Mol. Biol. 36:366–374.

Ogle, D. H. 2017. FSA: Fisheries Stock Analysis. R package version 0.8.13. https://CRAN.R-project.org/package=FSA.

Oksanen, J., F. Guillaume Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. Minchin, R. O'Hara, G. Simpson, P. Solymos, H. Stevens, E. Szoecs, and H. Wagner. 2017. vegan: Community Ecology Package. R package version 2.4-2.

Okumura, Y., Y. Ozeki, T. Itoh, S. Ohta, and H. Ômura. 2016. Volatile terpenoids from male wings lacking scent scales in *Anthocharis scolymus* (Lepidoptera: Pieridae). Appl. Entomol. Zool. 51:385–392.

Ômura, H., and K. Honda. 2005. Chemical composition of volatile substances from adults of the swallowtail, *Papilio polytes* (Lepidoptera: Papilionidae). Appl. Entomol. Zool. 40:421–427.

Ômura, H., K. Honda, and N. Hayashi. 2001. Identification of odoriferous compounds from adults of a swallowtail butterfly, *Papilio machaon* (Lepidoptera: Papilionidae). Z. Naturforschung C J. Biosci. 56:1126–1134.

Ômura, H., and S. Yotsuzuka. 2015. Male-specific epicuticular compounds of the sulfur butterfly *Colias erate poliographus* (Lepidoptera: Pieridae). Appl. Entomol. Zool. 50:191–199.

Otto, T. D., G. P. Dillon, W. S. Degrave, and M. Berriman. 2011. RATT: Rapid Annotation Transfer Tool. Nucleic Acids Res 39:e57.

Papke, R. S., D. J. Kemp, and R. L. Rutowski. 2007. Multimodal signalling: structural ultraviolet reflectance predicts male mating success better than pheromones in the butterfly *Colias eurytheme* L. (Pieridae). Anim. Behav. 73:47–54.

Paradis, E., and K. Schliep. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, doi: 10.1093/bioinformatics/bty633.

Partan, S. R. 2013. Ten unanswered questions in multimodal communication. Behav. Ecol. Sociobiol. 67:1523–1539.

Paten, B., D. Earl, N. Nguyen, M. Diekhans, D. Zerbino, and D. Haussler. 2011b. Cactus: Algorithms for genome multiple sequence alignment. Genome Res. 21:1512–1528.

Paten, B., M. Diekhans, D. Earl, J. S. John, J. Ma, B. Suh, and D. Haussler. 2011a. Cactus graphs for genome comparisons. J. Comput. Biol. 18:469–481.

Pfennig, K. S. 1998. The evolution of mate choice and the potential for conflict between species and mate–quality recognition. Proc. R. Soc. Lond. B Biol. Sci. 265:1743–1748.

Phelan, P. L., and T. C. Baker. 1986. Male-size-related courtship success and intersexual selection in the tobacco moth, *Ephestia elutella*. Experientia 42:1291–1293.

Phelan, P. L., and T. C. Baker. 1987. Evolution of male pheromones in moths: reproductive isolation through sexual selection? Science 235:205–207.

Phelan, P. L., and T. C. Baker. 1990. Comparative study of courtship in twelve phycitine moths (Lepidoptera: Pyralidae). J. Insect Behav. 3:303–326.

Picelli, S., A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, and R. Sandberg. 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 24:2033–2040.

Pinharanda, A., M. Rousselle, S. H. Martin, J. J. Hanly, J. W. Davey, S. Kumar, N. Galtier, and C. D. Jiggins. 2019. Sexually dimorphic gene expression and transcriptome evolution provide mixed evidence for a fast-Z effect in *Heliconius*. J. Evol. Biol. 32:194–204.

Pinzari, M., M. Santonico, G. Pennazza, E. Martinelli, R. Capuano, R. Paolesse, M. D. Rao, A. D'Amico, D. Cesaroni, V. Sbordoni, and C. D. Natale. 2018. Chemically mediated species recognition in two sympatric Grayling butterflies: *Hipparchia fagi* and *Hipparchia hermione* (Lepidoptera: Nymphalidae, Satyrinae). PLOS ONE 13:e0199997.

Pliske, T. E., and T. Eisner. 1969. Sex Pheromone of the Queen Butterfly: Biology. Science 164:1170–1172.

Poshyvailo, L., E. von Lieres, and S. Kondrat. 2017. Does metabolite channeling accelerate enzyme-catalyzed cascade reactions? PLoS ONE 12.

Poulton, E. B. 1925. The scents emitted by the neotropical butterfly *Heliconius erato hydarus* Hew. Trans. R. Entomol. Soc. Lond. 73:xxxvii–xlii.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2009. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 26:1641–1650.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2 – Approximately maximum-likelihood trees for large alignments. PLOS ONE 5:e9490.

Priyam, A., B. J. Woodcroft, V. Rai, A. Munagala, I. Moghul, F. Ter, M. A. Gibbins, H. Moon, G. Leonard, W. Rumpf, and Y. Wurm. 2015. Sequenceserver: a modern graphical user interface for custom BLAST databases. bioRxiv 033142.

Ptacek, M. 2000. The role of mating preferences in shaping interspecific divergence in mating signals in vertebrates. Behav. Processes 51:111–134.

Quental, T. B., M. M. Patten, and N. E. Pierce. 2007. Host plant specialization driven by sexual selection. Am. Nat. 169:830–836.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rantala, M. J., R. Kortet, J. S. Kotiaho, A. Vainikka, and J. Suhonen. 2003. Condition dependence of pheromones and immune function in the grain beetle *Tenebrio molitor*. Funct. Ecol. 17:534–540.

Rastas, P. 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. Bioinforma. Oxf. Engl. 33:3726–3732.

Rastogi, S., and D. A. Liberles. 2005. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. BMC Evol. Biol. 5:28.

Ready, J. S., I. Sampaio, H. Schneider, C. Vinson, T. D. Santos, and G. F. Turner. 2006. Colour forms of Amazonian cichlid fish represent reproductively isolated species. J. Evol. Biol. 19:1139–1148.

Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16:276–277.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Robinson, M. D., and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11:R25.

Roelofs, W. L., W. Liu, G. Hao, H. Jiao, A. P. Rooney, and C. E. Linn. 2002. Evolution of moth sex pheromones via ancestral genes. Proc. Natl. Acad. Sci. U. S. A. 99:13621–13626.

Roelofs, W. L., and A. P. Rooney. 2003. Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. Proc. Natl. Acad. Sci. U. S. A. 100:9179–9184.

Rojas, B., E. Burdfield-Steel, C. De Pasqual, S. Gordon, L. Hernández, J. Mappes, O. Nokelainen, K. Rönkä, and C. Lindstedt. 2018. Multimodal Aposematic Signals and Their Emerging Role in Mate Attraction. Front. Ecol. Evol. 6:93.

Rojas, B., E. Burdfield-Steel, H. Pakkanen, K. Suisto, M. Maczka, S. Schulz, and J. Mappes. 2017. How to fight multiple enemies: target-specific chemical defences in an aposematic moth. Proc. Biol. Sci. 284:20171424.

Rojas, B., J. Mappes, and E. Burdfield-Steel. 2019. Multiple modalities in insect warning displays have additive effects against wild avian predators. Behav. Ecol. Sociobiol. 73:37.

Rothschild, M. 1961. Defensive Odours and Mullerian Mimicry Among Insects. Trans. R. Entomol. Soc. Lond. 113:101–123.

Rothschild, M., B. P. Moore, and W. V. Brown. 1984. Pyrazines as warning odour components in the Monarch butterfly, *Danaus plexippus*, and in moths of the genera *Zygaena* and *Amata* (Lepidoptera). Biol. J. Linn. Soc. 23:375–380.

Ruther, J., M. Matschke, L.-A. Garbe, and S. Steiner. 2009. Quantity matters: male sex pheromone signals mate quality in the parasitic wasp *Nasonia vitripennis*. Proc. R. Soc. B Biol. Sci. 276:3303–3310.

Ryan, M. J., A. S. Rand, and L. A. Weigt. 1996. Allozyme and Advertisement Call Variation in the Túngara Frog, Physalaemus Pustulosus. Evolution 50:2435–2453.

Ryan, M. J., and A. S. Rand. 1993. Species Recognition and Sexual Selection as a Unitary Problem in Animal Communication. Evolution 47:647–657.

Ryan, M. J., and M. A. Guerra. 2014. The mechanism of sound production in túngara frogs and its role in sexual selection and speciation. Curr. Opin. Neurobiol. 28:54–59.

Sakai, R., M. Fukuzawa, R. Nakano, S. Tatsuki, and Y. Ishikawa. 2009. Alternative suppression of transcription from two desaturase genes is the key for species-specific sex pheromone biosynthesis in two *Ostrinia* moths. Insect Biochem. Mol. Biol. 39:62–67.

Sánchez, A. P., C. Pardo-Diaz, J. Enciso-Romero, A. G. Muñoz, C. D. Jiggins, C. Salazar, and M. Linares. 2015. An introgressed wing pattern acts as a mating cue. Evol. Int. J. Org. Evol. 69:1619–1629.

Saveer, A. M., P. G. Becher, G. Birgersson, B. S. Hansson, P. Witzgall, and M. Bengtsson. 2014. Mate recognition and reproductive isolation in the sibling species *Spodoptera littoralis* and *Spodoptera litura*. Chem. Ecol. 2:18.

Schiestl, F. P. 2010. The evolution of floral scent and insect chemical communication. Ecol. Lett. 13:643–656.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9:671–675.

Schneider, D. 1992. 100 years of pheromone research. An essay on Lepidoptera. Naturwissenschaften 79:241–250.

Schulz, S. 2009. Alkaloid-derived male courtship pheromones. Pp. 145–153 in W. E. Conner, ed. Tiger Moths and Woolly Bears: Behaviour, Ecology, and Evolution of the Arctiidae. Oxford University Press, Oxford.

Schulz Stefan, Boppré Michael, and Vane-Wright R. I. 1993. Specific mixtures of secretions from male scent organs of African milkweed butterflies (Danainae). Philos. Trans. R. Soc. Lond. B. Biol. Sci. 342:161–181.

Schulz, S., C. Estrada, S. Yildizham, M. Boppré, and L. E. Gilbert. 2008. An antiaphrodisiac in *Heliconius melpomene* butterflies. J. Chem. Ecol. 34:82–93.

Schulz, S., S. Yildizhan, K. Stritzke, C. Estrada, and L. E. Gilbert. 2007. Macrolides from the scent glands of the tropical butterflies *Heliconius cydno* and *Heliconius pachinus*. Org. Biomol. Chem. 5:3434.

Seehausen, O., Y. Terai, I. S. Magalhaes, K. L. Carleton, H. D. J. Mrosso, R. Miyagi, I. van der Sluijs, M. V. Schneider, M. E. Maan, H. Tachida, H. Imai, and N. Okada. 2008. Speciation through sensory drive in cichlid fish. Nature 455:620–626.

Selz, O. M., M. E. R. Pierotti, M. E. Maan, C. Schmid, and O. Seehausen. 2014. Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. Behav. Ecol. 25:612–626.

Shelly, T. E., J. Edu, and E. Pahio. 2007. Condition-dependent mating success in male fruit flies: ingestion of a pheromone precursor compensates for a low-quality diet. J. Insect Behav. 20:347–365.

Sheppard, P. M., J. R. G. Turner, Brown K. S., Benson W. W., and Singer M. C. 1985. Genetics and the evolution of Muellerian mimicry in *Heliconius* butterflies. Philos. Trans. R. Soc. Lond. B Biol. Sci. 308:433–610.

Sherratt, T. N. 2008. The evolution of Müllerian mimicry. Naturwissenschaften 95:681–695.

Smadja, C. M., and R. K. Butlin. 2008. On the scent of speciation: the chemosensory system and its role in premating isolation. Heredity 102:77–97.

Smiley, J. 1978. Plant chemistry and the evolution of host specificity: new evidence from *Heliconius* and *Passiflora*. Science 201:745–747.

South, S. H., C. M. House, A. J. Moore, S. J. Simpson, and J. Hunt. 2011. Male cockroaches prefer a high carbohydrate diet that makes them more attractive to females: implications for the study of condition dependence. Evolution 65:1594–1606.

Southcott, L., and M. R. Kronforst. 2018. Female mate choice is a reproductive isolating barrier in *Heliconius* butterflies. Ethology 124:862–869.

Srygley, R. B., and P. Chai. 1990. Flight morphology of Neotropical butterflies: palatability and distribution of mass to the thorax and abdomen. Oecologia 84:491–499.

Stern, D. L. 2013. The genetic causes of convergent evolution. Nat. Rev. Genet. 14:751–764.

Stoeffler, M., T. S. Maier, T. Tolasch, and J. L. M. Steidle. 2007. Foreign-language Skills in Rove-Beetles? Evidence for Chemical Mimicry of Ant Alarm Pheromones in Myrmecophilous *Pella* Beetles (Coleoptera: Staphylinidae). J. Chem. Ecol. 33:1382–1392.

Sun, K., L. Luo, R. T. Kimball, X. Wei, L. Jin, T. Jiang, G. Li, and J. Feng. 2013. Geographic Variation in the Acoustic Traits of Greater Horseshoe Bats: Testing the

Importance of Drift and Ecological Selection in Evolutionary Processes. PLOS ONE 8:e70368.

Supple, M., R. Papa, B. Counterman, and W. O. McMillan. 2014. The genomics of an adaptive radiation: insights across the *Heliconius* speciation continuum. Adv. Exp. Med. Biol. 781:249–271.

Swain, M. T., I. J. Tsai, S. A. Assefa, C. Newbold, M. Berriman, and T. D. Otto. 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat Protoc 7:1260–1284.

Symonds, M. R. E., and M. A. Elgar. 2008. The evolution of pheromone diversity. Trends Ecol. Evol. 23:220–228.

Symonds, M. R. E., T. L. Johnson, and M. A. Elgar. 2012. Pheromone production, male abundance, body size, and the evolution of elaborate antennae in moths. Ecol. Evol. 2:227–246.

Takanashi, T., Y. Huang, K. R. Takahasi, S. Hoshizaki, S. Tatsuki, and Y. Ishikawa. 2005. Genetic analysis and population survey of sex pheromone variation in the adzuki bean borer moth, *Ostrinia scapulalis.* Biol. J. Linn. Soc. 84:143–160.

The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47:D506–D515.

Tholl, D. 2006. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. Curr. Opin. Plant Biol. 9:297–304.

Thurman, T. J., E. Brodie, E. Evans, and W. O. McMillan. 2018. Facultative pupal mating in *Heliconius erato*: Implications for mate choice, female preference, and speciation. Ecol. Evol. 8:1882–1889.

Thurmond, J., J. L. Goodman, V. B. Strelets, H. Attrill, L. S. Gramates, S. J. Marygold, B. B. Matthews, G. Millburn, G. Antonazzo, V. Trovisco, T. C. Kaufman, B. R. Calvi, N. Perrimon, S. R. Gelbart, J. Agapite, K. Broll, L. Crosby, G. dos Santos, D. Emmert, L. S. Gramates, K. Falls, V. Jenkins, B. Matthews, C. Sutherland, C. Tabone, P. Zhou, M.

Zytkovicz, N. Brown, G. Antonazzo, H. Attrill, P. Garapati, A. Holmes, A. Larkin, S. Marygold, G. Millburn, C. Pilgrim, V. Trovisco, P. Urbano, T. Kaufman, B. Calvi, B. Czoch, J. Goodman, V. Strelets, J. Thurmond, R. Cripps, and P. Baker. 2019. FlyBase 2.0: the next generation. Nucleic Acids Res. 47:D759–D765.

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7:562–578.

Trivers, R. L. 1972. Parental Investment and Sexual Selection. Pp. 136–179 in B. G. Campbell, ed. Sexual Selection and the Descent of Man. Aldine, Chicago.

Van Belleghem, S. M., P. Rastas, A. Papanicolaou, S. H. Martin, C. F. Arias, M. A. Supple, J. J. Hanly, J. Mallet, J. J. Lewis, H. M. Hines, M. Ruiz, C. Salazar, M. Linares, G. R. P. Moreira, C. D. Jiggins, B. A. Counterman, W. O. McMillan, and R. Papa. 2017. Complex modular architecture around a simple toolkit of wing pattern genes. Nat. Ecol. Evol. 1:52.

van Bergen, E., P. M. Brakefield, S. Heuskin, B. J. Zwaan, and C. M. Nieberding. 2013. The scent of inbreeding: a male sex pheromone betrays inbred males. Proc. R. Soc. Lond. B Biol. Sci. 280:20130102.

van Schooten, B., C. D. Jiggins, A. D. Briscoe, and R. Papa. 2016. Genome-wide analysis of ionotropic receptors provides insight into their evolution in *Heliconius* butterflies. BMC Genomics 17:254.

Vane-Wright, D. 1972. Scent organs of male butterflies. Rep. Br. Mus. Nat. Hist. 1969–1971:31–35, 4 pls.

Vane-Wright, R. I., and M. Boppré. 1993. Visual and Chemical Signalling in Butterflies: Functional and Phylogenetic Perspectives. Philos. Trans. R. Soc. Lond. B Biol. Sci. 340:197–205.

Vencl, F. V., K. Ottens, M. M. Dixon, S. Candler, X. E. Bernal, C. Estrada, and R. A. Page. 2016. Pyrazine emission by a tropical firefly: An example of chemical aposematism? Biotropica 48:645–655.

Vereecken, N. J., and J. N. McNeil. 2010. Cheaters and liars: chemical mimicry at its finest. Can. J. Zool. 88:725–752.

Visser, B., I. A. N. Dublon, S. Heuskin, F. Laval, P. Bacquet, G. Lognay, and C. M. Nieberding. 2018. Common practice solvent extraction does not reflect actual emission of a sex pheromone during butterfly courtship. Front. Ecol. Evol. 6:154.

Wagner, A. 1998. The fate of duplicated genes: loss or new function? BioEssays 20:785–788.

Walters, J. R., T. J. Hardcastle, and C. D. Jiggins. 2015. Sex chromosome dosage compensation in *Heliconius* butterflies: Global yet still incomplete? Genome Biol Evol 7:2545–2559.

Walters, J. R., C. Stafford, T. J. Hardcastle, and C. D. Jiggins. 2012. Evaluating female remating rates in light of spermatophore degradation in *Heliconius* butterflies: pupal-mating monandry versus adult-mating polyandry. Ecol. Entomol. 37:257–268.

Wang, H.-L., O. Brattström, P. M. Brakefield, W. Francke, and C. Löfstedt. 2014. Identification and Biosynthesis of Novel Male Specific Esters in the Wings of the Tropical Butterfly, *Bicyclus martius sanaos*. J. Chem. Ecol. 40:549–559.

Wang, H.-L., M. A. Liénard, C.-H. Zhao, C.-Z. Wang, and C. Löfstedt. 2010a. Neofunctionalization in an ancestral insect desaturase lineage led to rare Δ6 pheromone signals in the Chinese tussah silkworm. Insect Biochem. Mol. Biol. 40:742–751.

Wang, H.-L., G. P. Svensson, O. Rosenberg, M. Bengtsson, E. V. Jirle, and C. Löfstedt. 2010b. Identification of the Sex Pheromone of the Spruce Seed Moth, *Cydia strobilella* L. J. Chem. Ecol. 36:305–313.

Wang, K. C., and S. Ohnuma. 2000. Isoprenyl diphosphate synthases. Biochim. Biophys. Acta 1529:33–48.

Wang, Y., U. Naumann, S. T. Wright, and D. I. Warton. 2012. mvabund– an R package for model-based analysis of multivariate abundance data. Methods Ecol. Evol. 3:471–474.

Warton, D. I., S. T. Wright, and Y. Wang. 2012. Distance-based multivariate analyses confound location and dispersion effects. Methods Ecol. Evol. 3:89–101.

Weber, M. G., L. Mitko, T. Eltz, and S. R. Ramírez. 2016. Macroevolution of perfume signalling in orchid bees. Ecol. Lett. 19:1314–1323.

Wegehaupt, A. K., and W. E. Wagner. 2017. Females can solve the problem of low signal reliability by assessing multiple male traits. Biol. Lett. 13:20170386.

Wertheim, B., E.-J. A. van Baalen, M. Dicke, and L. E. M. Vet. 2005. Pheromone-mediated aggregation in nonsocial arthropods: an evolutionary ecological perspective. Annu. Rev. Entomol. 50:321–346.

Wicker-Thomas, C. 2011. Evolution of insect pheromones and their role in reproductive isolation and speciation. Ann. Société Entomol. Fr. NS 47:55–62.

Wickham, H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wiklund, C., and J. Forsberg. 1986. Courtship and male discrimination between virgin and mated females in the orange tip butterfly *Anthocharis cardamines*. Anim. Behav. 34:328–332.

Wong, B. 2011. Points of view: Color blindness. Nat. Methods 8:441.

Wyatt, T. D. 2003. Pheromones and Animal Behaviour: Communication by Smell and Taste. Cambridge University Press, Cambridge.

Wyatt, T. D. 2014. Pheromones and Animal Behavior: Chemical Signals and Signatures. Cambridge University Press, Cambridge.

Yildizhan, S., J. van Loon, A. Sramkova, M. Ayasse, C. Arsene, C. ten Broeke, and S. Schulz. 2009. Aphrodisiac pheromones from the wings of the small cabbage white and large cabbage white butterflies, *Pieris rapae* and *Pieris brassicae*. ChemBioChem 10:1666–1677.

Zahavi, A. 1975. Mate selection—A selection for a handicap. J. Theor. Biol. 53:205–214.

Zahavi, A. 1977. The cost of honesty: further remarks on the handicap principle. J. Theor. Biol. 67:603–605.

Zhu, B.-Q., J. Cai, Z.-Q. Wang, X.-Q. Xu, C.-Q. Duan, and Q.-H. Pan. 2014. Identification of a plastid-localized bifunctional nerolidol/linalool synthase in relation to linalool biosynthesis in young grape berries. Int. J. Mol. Sci. 15:21992–22010.

Zimmermann, Y., D. W. Roubik, and T. Eltz. 2006. Species-specific attraction to pheromonal analogues in orchid bees. Behav. Ecol. Sociobiol. 60:833.

Appendix

A major locus controls a biologically active pheromone component in *Heliconius melpomene*

Running title: Genetics of pheromones in *Heliconius*

Kelsey J.R.P. Byers[1,2,7], Kathy Darragh[1,2,7], Jamie Musgrove[2], Diana Abondano Almeida[2,†], Sylvia Fernanda Garza[2,‡], Ian A. Warren[1], Pasi M. Rastas[3], Marek Kučka[4], Yingguang Frank Chan[4], Richard M. Merrill[5], Stefan Schulz[6], W. Owen McMillan[2], Chris D. Jiggins[1,2,8]

[1] Department of Zoology, University of Cambridge, Cambridge, United Kingdom

[2] Smithsonian Tropical Research Institute, Panama, Panama

[3] Institute of Biotechnology, University of Helsinki, Helsinki, Finland

[4] Friedrich Miescher Laboratory of the Max Planck Society, Tuebingen, Germany

[5] Division of Evolutionary Biology, Ludwig-Maximilians-Universität München, Munich, Germany

[6] Institute of Organic Chemistry, Department of Life Sciences, Technische Universität Braunschweig, Braunschweig, Germany

[7] Kelsey J.R.P. Byers and Kathy Darragh should be considered joint first author

[8] To whom correspondence should be addressed: c.jiggins@zoo.cam.ac.uk

[†] Present address: Institute for Ecology, Evolution and Diversity, Goethe Universität, Frankfurt, Germany

[‡] Present address: Department of Collective Behaviour, Max Planck Institute of Animal Behaviour, Konstanz, Germany & Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Konstanz, Germany

Abstract

Pheromones are important for courtship and mate choice in many insects, but we know relatively little of their role in butterflies. The butterfly *Heliconius melpomene* uses a complex blend of wing androconial compounds during courtship. Electroantennography in *H. melpomene* and its close relative *H. cydno* showed that responses to androconial extracts were not species-specific. Females of both species responded more strongly to the *H. cydno* extract, suggesting conservation of peripheral nervous system elements across the two species. Individual blend components provoked little to no response, with the exception of octadecanal, a major component of the *H. melpomene* blend. Supplementing octadecanal on the wings of octadecanal-rich *H. melpomene* males led to an increase in the time until mating, demonstrating the bioactivity of octadecanal in *Heliconius*. Using quantitative trait locus (QTL) mapping, we identified a single locus on chromosome 20 responsible for 41% of the parental species' difference in octadecanal production. This QTL does not overlap with any of the major wing color or mate choice loci, nor does it overlap with known regions of elevated or reduced F_{ST}. A set of 16 candidate fatty acid biosynthesis genes lies underneath the QTL.

Keywords: pheromones; electroantennography; behavior; quantitative trait locus mapping; *Heliconius*

Introduction

Chemical communication is the oldest form of sensory communication, and plays a fundamental role in the ecology of organisms across the tree of life. In terms of reproductive behavior, chemical communication is involved in premating isolation in a variety of systems from orchids (Peakall et al. 2010) to *Drosophila* (Shahandeh et al. 2017) to cichlids (Plenderleith et al. 2005), highlighting its importance as a mediator of speciation and diversification (Smadja & Butlin 2009). Of particular interest is the evolution of pheromones, chemical compounds that mediate intraspecies interactions. In particular, signaling (production and emission of pheromone compounds) and receiving (reception and interpretation of chemical signals) components are predicted to evolve in concert to maintain reproductive isolation between closely related species (Smadja & Butlin 2009). Nonetheless, despite its ubiquity, chemical communication has been less

well studied than *e.g.* visual communication. The result is that outside of a limited set of (relatively well-studied) examples we know relatively little about the role that chemical signals play in reproductive isolation and evolution. Nonetheless, recent technical advances (e.g. later-generation gas chromatography-coupled mass spectrometry and gas chromatography-coupled electroantennographic detection) mean that chemical communication is increasingly accessible for evolutionary studies.

Studies of pheromones have been most widespread in insects, in particular in the order Lepidoptera and in *Drosophila* (Smadja & Butlin 2009). In Lepidoptera, work on pheromones has largely focused on long-range female pheromones of nocturnal moths (often due to their economic importance), where pheromone divergence commonly contributes to speciation and relatively simple structural variations in pheromones are known to produce drastic differences in mate attraction. For example, different populations of the corn borer moth *Ostrinia nubialis* exhibit a simple *cis* to *trans* switch in the pheromone 11-tetradecenyl acetate (Kochansky et al. 1975, Lassance et al. 2010) that leads to partial reproductive isolation (Dopman et al. 2010). Sex pheromones have been less well studied in day-flying butterflies, where visual signaling is often assumed to play a more dominant role in mate choice (Vane-Wright and Boppré 1993, Löfstedt et al. 2016). However, male butterflies also emit close-range pheromone bouquets which may act in concert with other wing pattern and behavioral cues (Mérot et al. 2015), and are important in mate choice (Darragh et al. 2017) as well as decreasing heterospecific mating (Mérot et al. 2015).

Despite the potential importance of pheromones in butterflies, aphrodisiac pheromones have so far been identified in only eight butterfly species (Meinwald et al. 1969; Pliske and Eisner 1969; Grula et al. 1980; Nishida et al. 1996; Schulz and Nishida 1996; Andersson et al. 2007; Nieberding et al. 2008; Yildizhan et al. 2009). This is in stark contrast to the approximately two thousand species of moths where female pheromones or attractants are known (Löfstedt et al. 2016). There is a similar absence of knowledge about the genetic basis of variation in pheromone production (but see Liénard et al. 2014). As in other diurnal butterflies, the chemical bouquets of the genus *Heliconius* (Darragh et al. 2017; Mann et al. 2017), are complex, both in identity and quantity of compounds, though just a few individual compounds may be biologically active

pheromones. Although studying variation in pheromones within and across species can point to potential candidates (e.g. Darragh et al. 2019b), identifying which components of these complex chemical bouquets are responsible for pheromonal communication is a considerable challenge. This is particularly true as even minor compounds can have major effects (McCormick et al. 2014; Chen et al. 2018). Determining pheromone bioactivity requires screening compounds via physiological activity followed by behavioral verification, and in many cases pheromone bouquet composition is known but the bioactive components remain unidentified. In addition, behavioral outcomes may differ despite similar responses in the peripheral nervous system (Chen & Fadamiro 2007; Seeholzer et al. 2018).

Here we take advantage of two closely related *Heliconius* butterfly species, *H. melpomene* and *H. cydno*, to further our knowledge of the ecology, evolution and genetics of male lepidopteran pheromones. The two species diverged about 2.1 million years ago (Arias et al 2014; Kozak et al. 2015), and are strongly reproductively isolated (Jiggins 2017). Over the past decade there has been considerable research into the genomic architecture of differences in wing pattern and male mate preference (Jiggins 2017; Merrill et al. 2019). Surprisingly, both wing color and male mating preferences between these species have a relatively simple genetic basis with a large proportion of the difference between parental forms being controlled by a small handful of loci of large effect (Naisbit 2003; Jiggins 2017; Merrill et al. 2019). This has important implications for speciation, as theory predicts that large effect loci contribute to speciation in the face of gene flow (Via 2012). Similarly, tight physical linkage between loci that contribute to isolating barriers will facilitate speciation (Felsenstein 1981; Merrill et al. 2010; Smadja and Butlin 2011), and there is evidence for tight linkage of a gene controlling wing pattern (*optix*) and a major effect QTL underlying divergent male preference behaviors (Merrill et al. 2019). Pheromonal differences between the species might also be expected to be under control by major effect loci, as has been seen in a variety of moth species (Groot et al. 2016, Haynes et al. 2016). The extent to which loci underlying pheromone production overlap with wing pattern and mate choice loci is unclear, but given the existing linkage between wing pattern and male mate choice loci, we might predict additional linkage of pheromone production with wing pattern loci.

Recent research has demonstrated the importance of male pheromones in mating success (Darragh et al. 2017). To better characterize male butterfly pheromone production and the role it plays in mating and species recognition, we comprehensively analyzed pheromone bouquets of *Heliconius melpomene* and *H. cydno* butterflies to determine the most important bioactive compounds. We carried out electrophysiology and behavioral experiments and identified octadecanal as a biologically active pheromone component. To better understand the genetic basis of octadecanal production and determine the location of loci responsible for the production of octadecanal relative to loci involved in wing color pattern and male mating preference, we mapped loci responsible for differences in the level of octadecanal synthesis between the two species. We found that a single locus on chromosome 20 explains 41% of the parental species differences in this pheromone, with no linkage between this locus and known color pattern and male mate choice loci.

Materials and Methods

Butterflies

Stocks from central Panama of *Heliconius melpomene rosina* and *H. cydno chioneus* (hereafter *H. melpomene* and *H. cydno*) were used for all experiments (Darragh et al. 2017). Butterflies were reared in insectaries at the Smithsonian Tropical Research Institute, Gamboa, Panama under ambient temperature and light conditions. Eggs were collected from these breeding stocks and the resulting larvae fed on *Passiflora platyloba var. williamsi*, *P. biflora*, *P. menispermifolia*, and *P. vitifolia* until pupation. Data from (Darragh et al. 2019a) indicate that larval diet does not affect the major compounds found in *H. melpomene*, suggesting that this dietary variation is unlikely to affect results. Newly-eclosed adult butterflies were separated by sex to ensure virgin status and supplied with flowers from *Psiguria warscewiczii*, *Psiguria triphylla*, *Gurania eriantha*, *Psychotria poeppigiana*, *Stachytarpheta mutabilis*, and *Lantana* sp. (most likely *L. camara*) as pollen sources, as well as a ~20% sucrose solution. All experiments used virgin butterflies. For assessment of *H. cydno* wing bouquets, male butterflies were between 10-12 days post eclosion and had fed exclusively on *Passiflora platyloba var. williamsi*. For electrophysiology, female butterflies were between 1 and 20 days post eclosion, and

males between 10-20 days post eclosion to ensure male sexual maturity (Darragh et al. 2017). For behavior, female butterflies were used the day after eclosion and males were between 10 and 59 days post eclosion. Natural wing extracts of both species were extracted from males 10-12 days post eclosion as described in (Darragh et al. 2017) using dichloromethane plus 1 ng/µL 2-tetradecyl acetate (hereafter "DCM+IS") and concentrated approximately 10x prior to use under still room air. All samples were stored at -20°C before use.

Identification and quantification of androconial compounds

To identify species-specific compounds among our two species, the chemical composition of the *H. cydno* androconial bouquet was investigated in samples from 26 adult male *H. cydno* and compared with 31 adult male *H. melpomene*, the latter including samples previously analyzed in (Darragh et al. 2017; Darragh et al. 2019a), all collected as above. Samples were assessed using gas chromatography-mass spectrometry (GC-MS) with an Agilent 7890B gas chromatograph coupled with an Agilent 5977 mass spectrometer with electron ionization (Agilent Technologies, California, USA). The GC utilized an Agilent HP-5MS capillary column (30m length, 0.25mm inner diameter), helium carrier gas at 1.2 mL/minute, and an Agilent ALS 7693 autosampler. Injection was splitless with an inlet temperature of 250°C. The temperature ramp was isothermal at 50°C for 5 minutes, then increased at 5°C/minute to 320°C and was then isothermal for 5 minutes. Samples were identified using a custom MS library and quantified by comparison with the internal standard.

In line with (Darragh et al. 2017), wings from eight male *H. cydno* were dissected into four regions: hindwing androconia, forewing overlap region, hindwing rest-of-wing, and forewing rest-of-wing, and all extracted identically after dissection. Wing region area was quantified by photographing the wings before dissection and measuring the total pixel area of each wing region in the GNU Image Manipulation Program v.2.8.20 (GIMP Development Team), with the pixel-mm conversion via measurement of a ruler in the photograph. Quantified compounds in each wing region for each individual were scaled by the area of the relevant wing region in that individual.

Chemicals

Syringaldehyde (3,5-dimethoxy-4-hydroxybenzaldehyde), 1-octadecanol, and henicosane were obtained commercially (Sigma-Aldrich). The aldehydes octadecanal, (*Z*)-11-icosenal, and (*Z*)-13-docosenal were obtained from the respective alcohols 1-octadecanol, (*Z*)-11-icosen-1-ol, and (*Z*)-13-docosen-1-ol by oxidation with iodoxybenzoic acid (IBX) in ethyl acetate according to (More and Finney 2002). The required alcohols (*Z*)-11-icosen-1-ol and (*Z*)-13-docosen-1-ol were in turn obtained by lithium aluminum hydride reduction from commercially available (*Z*)-11-icosenoic acid and methyl (*Z*)-13-docosenoate (Larodan) according to (Cha and Brown 1993). The seven target compounds (see Figures S4 and S5 for structures and reaction scheme) were chosen due to their quantitative dominance in the chemical profiles of *H. melpomene* and *H. cydno*. The solvent for all synthesized compounds was hexane, with the exception of the polar syringaldehyde, which was diluted in a 1:10 mixture of dichloromethane and hexane.

Synthetic blends of *H. melpomene* and *H. cydno* male wing bouquets were prepared from these synthesized compounds. Synthetic *H. melpomene* contained 23.2 ng/µL syringaldehyde, 23.3 ng/µL octadecanal, 6.9 ng/µL 1-octadecanol, 4.7 ng/µL (*Z*)-11-icosenal, 20.3 ng/µL (*Z*)-11-icosenol, and 4.8 ng/µL (*Z*)-13-docosenal. Synthetic *H. cydno* contained 47.0 ng/µL syringaldehyde and 93.3 ng/µL henicosane. Floral direct extractions from *Lantana* sp. (most likely *L. camara*) growing wild in Gamboa, Panama were used as a positive control. Single umbels were removed from plants at dawn and placed in a scintillation vial to which 400µL of DCM+IS was added. After 1 hour, the DCM+IS was removed to a glass vial and kept at -20°C before use.

Electroantennography

Electrophysiological preparations were assembled as follows: antennae were cut from the head of a virgin butterfly using fine scissors and the final 6.5 segments (approximately 1.5mm, avoiding cutting at the segment boundary) cut off with a scalpel. Both antennae were then placed in parallel across an antenna fork (Syntech, Buchenbach, Germany) and held in place using electrode gel (Spectra 360 Electrode Gel, Parker Laboratories Inc., Fairfield, NJ, USA). The antenna fork was mounted on a Syntech EAG CombiProbe with 10x internal gain, and signal from this was routed through a Syntech IDAC4. EAG waveforms recorded using Syntech GcEad/2014 software. Stimulus pulses

were delivered using a Syntech CS-55 Stimulus Controller with foot pedal trigger. Both continuous unhumified room air and stimulus pulses were delivered to the preparation at 1.5 liters/min through a tube of approximately 8mm inner diameter. The stimulus pulses were delivered in triplets of 0.5 seconds each, separated by 5 seconds, with triplets initiated every 30 seconds. Stimulus delivery used odor cartridges assembled from Pasteur pipettes with a strip of filter paper plugged with cotton when not in use; each stimulus cartridge was 'charged' with 10uL of stimulus solution for each experiment. Each antennal preparation was used only once.

Two sets of stimuli were delivered: a species comparison set and a synthetic compound set. Both sets used air (nothing added to filter paper), hexane, and DCM+IS as negative controls and *Lantana* extract as a positive control. The species comparison set included male wing extracts from *H. melpomene* and *H. cydno* ("Mnat" and "Cnat" respectively) and synthetic blends representing the two species ("Msyn" and Csyn"). The synthetic compound set included air, hexane, DCM+IS, the conspecific male wing extract, the conspecific synthetic blend, and the seven synthetic compounds. Presentation order was randomized before each experiment. Species comparison experiments consisted of sixteen pulses each of the seven stimuli, interspersed with five pulses of *Lantana* extract at the start, between every three stimuli, and at the end. Synthetic compounds were similar, with eleven pulses of each of the twelve stimuli, interspersed with four pulses of *Lantana* at the start, between every four stimuli, and at the end. For analysis, the first triplet of each stimulus set was removed, leaving fifteen and ten pulses respectively. At least ten female and five male butterflies of each species were used with each experiment.

Onset time and amplitude of EAG responses (i.e. the magnitude of the decrease from the baseline signal, barred lines in Figure S6) were marked using GcEad/2014. To control for antennal signal degradation over time, a time-dependent correction factor was calculated using linear interpolation between each *Lantana* set and this applied to the absolute value of the EAG response amplitude. These corrected amplitudes were then scaled to the amplitude of the initial *Lantana* set to partially control for differences between preparations.

Analysis of antennal adaptation

Short-term adaptation (STA), as defined in (Zufall and Leinders-Zufall 2000), is adaptation occurring only over a very brief window from the initial stimulus that then resolves quickly with no further stimulation (e.g. the 10 second interval given for salamanders in the reference). By contrast, long-term or long-lasting adaptation (LTA) are defined in the same source as persisting over an extended period of time, up to several minutes in vertebrates and insects (Stengl 2010), with recovery of response upon presentation of a different stimulus. Within our electrophysiological data set, there is the potential to measure both types of adaptation; since we corrected for preparation degradation over time, we should be able to measure true LTA and STA. STA, if present, should be evident within an individual triplet, as the stimuli within a triplet are separated by 5 seconds (and thus STA is likely to persist within a triplet), whereas LTA should be evident across an individual stimulus set, as these lasted 5.5-8 minutes with maximum intervals of 30 seconds between stimuli, insufficient for LTA to be abolished if we assume similar mechanisms as vertebrates. We assessed antennal responses for LTA by pooling all triplets within a stimulus-species-sex combination. We pooled these data within stimulus set types (species comparison and synthetic compound), treating butterfly preparation identity as a random effect. For each stimulus, the change in antennal response was assessed over the time since initial presentation of the stimulus. STA was assessed by looking for significant changes in the residuals from this analysis between members of a triplet.

Behavior

To test the potential role of octadecanal in *H. melpomene* female mate choice, behavioral experiments were conducted in insectaries at STRI, Gamboa, Panama, between April and July 2018. One day old virgin females were presented with both an octadecanal treated and a control *H. melpomene* male for two hours. Males were at least ten days old and were selected based on similarity of size, wing-wear and age, with

treatment being allocated randomly by coin flip. Either 25 µl octadecanal solution (140 ng/µl octadecanal in hexane, thus adding 3500ng to the existing average 773.4ng for approximately a 5.5x dose) or 25 µl pure hexane (both evaporated down to a smaller volume of approximately 10 µl under room air) was applied to the hindwing androconial region of each male, and males were then allowed 30 minutes to settle before beginning the two hour experiment period. Experiments began at or close to 9am, with observations being made every 15 minutes or until mating occurred. *Heliconius melpomene* was chosen for these experiments as an adequate number of individuals were available, although the less easily reared (and thus less available) *H. cydno* might have provided a clearer picture of the role of octadecanal in premating reproductive isolation.

To test the persistence of the octadecanal treatment on the wings of live butterflies, a separate set of *H. melpomene* males was treated as above with either hexane or octadecanal. Separate males were sampled at 30 minutes post treatment and two hours post treatment by extraction of the forewing overlap region (Darragh et al. 2017) and the hindwing androconia in DCM+IS as above, with two males per treatment-time combination. Octadecanal was then measured using GC-MS as above and quantified by peak area comparison with the 2-tetradecyl acetate internal standard.

Quantitative trait locus mapping for octadecanal production

To map the genetic basis for octadecanal production in *H. melpomene*, we took advantage of the fact that *H. cydno* produces little to no octadecanal. Bidirectional F_1 crosses between the two species revealed that the *H. cydno* phenotype (low to no octadecanal) is dominant over the high octadecanal production found in *H. melpomene*, so we constructed backcross families by crossing F_1 males to female *H. melpomene* from our existing stocks. A total of ten families (nine with a female *H. melpomene* grandparent and one with a female *H. cydno* grandparent) were constructed, with each offspring representing a single recombination event from the F_1 father. We constructed backcrosses to *H. cydno* (127 individuals from 15 families) in addition, as some segregation was seen in this backcross direction as well. Butterflies were reared and wing extracts collected and analyzed from male offspring as described above, except that all

larvae were reared on *Passiflora platyloba var. williamsi*. Bodies of male offspring were collected into dimethyl sulfoxide (DMSO) for later library preparation. The Castle-Wright estimators for octadecanal and octadecanol production were calculated using the phenotypic variance of the backcross individuals as the estimated segregation variance (Jones 2001).

Qiagen DNeasy kits (Qiagen) were used for DNA extraction. Individuals were genotyped either by RAD-sequencing as previously described (Davey et al. 2017; Merrill et al. 2019), or low-coverage whole genome sequencing using nextera-based libraries. For the nextera-based libraries a secondary purification was performed using magnetic SpeedBeads™ (Sigma) dissolved in 0.44mM PEG8000, 2.5M NaCl, 1mM Tris-Cl pH=8, and 0.1mM EDTA pH=8.0. High-throughput sequencing libraries were generated using a method based on the Nextera DNA Library Prep (Illumina, Inc.) with purified Tn5 transposase (Picelli et al. 2014). Sample barcoding was performed using PCR extension with an i7-index primer (N701–N783) and the N501 i5-index primer. Libraries were purified and size selected using the same beads as above. Pooled libraries were sequenced by HiSeq 3000 (Illumina) by BGI (China).

Linkage mapping was conducted using standard Lep-MAP3(LM3) pipeline (Rastas 2017) . First, individual fastq files were mapped to the melpomene reference genome using BWA MEM (Li 2011) and then sorted bams were created using SAMtools (Li and Durbin 2011). The input genotype likelihoods were constructed by SAMtools mpileup and pileupParser2+pileup2posterior from LM3. The pedigree of individuals was validated and corrected using IBD (identity-by-descent) module and the sex of individuals was validated and corrected according to the coverage on the Z chromosome and autosomes using SAMtools depth. Then, ParentCall2 (parameter ZLimit=2) and Filtering2 (dataTolerance=0.001) modules were called on the input data and a random subset of 25% of markers (to speed up analysis) was used for the subsequent steps.

Initial linkage groups (chrX.map) and marker orders (orderX.txt) were constructed based on the *H. melpomene* genome for each of 21 chromosomes. SeparateChromosomes2 was run on each of these groups with the default parameters (lodLimit=10) except for map=chrX.map (for X=1..21). Finally, OrderMarkers2 was run on

each chromosome in the constructed order, with parameter scale=0.05, recombination2=0, evaluateOrder=orderX.txt and map=result_from_SeparateChromsomes2.chrX.txt. Another evaluation was done with data in the grandparental phase (additional parameter grandparentPhase=1). The phased data of these orders were matched using phasematch script (LM3) and obtaining all markers from the first evaluation in the grandparental phase. This obtained result was used as the final map.

Map construction resulted in the retention of 447,818 SNP markers across 89 and 127 individuals with phenotype data in backcrosses to *H. melpomene* and *H. cydno* respectively. To facilitate computation, markers were thinned evenly by a factor of ten, resulting in 44,782 markers with no missing data. Octadecanal production was log-transformed to obtain normality, then regressed against marker position using the R/qtl2 R library (Broman et al. 2018). Significance thresholds were obtained by permutation testing in R/qtl2 with 1000 permutations, and QTL confidence intervals obtained using the bayes_int command. To account for the family structure present in our QTL mapping populations, we additionally included a kinship matrix calculated by R/qtl2 using the LOCO (leave one chromosome out) method in the marker regression and recalculated significance thresholds and confidence intervals with the kinship term included. Percent variance explained was calculated as the difference in phenotype means of individuals of each genotype divided by the difference in the parental phenotype. Since the genetic linkage map was based on whole genome data, we were able to obtain physical positions of QTL confidence interval endpoints. The physical positions of the kinship-included confidence interval were used to query Lepbase (Challis et al. 2016) for potential candidate genes from the *H. melpomene* genome. To identify putative functions for each potential candidate, protein sequences from Lepbase were searched against the nr (non-redundant) protein database using BLASTp (Altschul et al. 1990). For each candidate with a promising functional annotation, exons were pulled out of the *H. cydno* genome (Pessoa Pinharanda 2017) and aligned to the *H. melpomene* genes using BLASTn with each *H. melpomene* exon to search for SNPs between the two species. Selection was tested at the sequence level using codeml (Yang 2007) in pairwise mode with the F3X4 codon frequency model.

Statistical analysis

Differences in individual compounds between *H. melpomene* and *H. cydno* were assessed using a Welch's *t*-test. Wing region differences in *H. cydno* were assessed for each individual compound found in at least four of eight samples of at least one wing region using a linear mixed model with wing region as a fixed effect and butterfly identity as a random effect using the package nlme v.3.1.137 (Pinheiro et al. 2018). Statistical tests for these models were assessed with the Anova function in the package car v.3.0.0 (Fox and Weisberg 2011), and comparisons between wing regions were performed using the emmeans package v.1.3.1 (Lenth 2018).

For electroantennography, species comparison sets and synthetic compound sets were analyzed separately. Corrected and scaled EAG responses (for each experiment within sexes and species) were compared between stimuli with a linear mixed model with stimulus as a fixed effect, butterfly preparation identity as a random effect, and the interaction of stimulus and preparation as a random effect using nlme as above. Statistical tests for these models were assessed with the Anova and emmeans (version 1.1.3) functions as above.

Long-term adaptation was assessed using robust linear mixed models with the package robustlmm v.2.2.1 (Koller 2016), with corrected but unscaled amplitude as the response variable, time since initial presentation of the stimulus as a fixed variable, and butterfly preparation as a random variable. Responses were pooled across samples within a species-sex combination and considered for each stimulus separately. As robustlmm does not provide p-values, confidence intervals were used to assess significance and difference between sample-species combinations. Short-term adaptation was assessed using the residuals from the same regression. Differences in the residuals between triplets within a triplet set were tested using a one-sample t-test with a hypothesized mean value of zero (i.e. no difference between residuals), performed on the subtractive difference between the residuals of the third (last) and first triplets.

Female mate choice was assessed using a binomial test, and treatment differences in time until mating were assessed with a two-sided t-test assuming unequal variances.

Octadecanal persistence was not assessed statistically due to the small sample size. All statistical tests were performed in R v.3.5.0, v.3.5.1, or v.3.5.2 (R Core Team 2013).

Results

Androconial chemical bouquets of *Heliconius melpomene* and *H. cydno*

In order to identify candidate pheromone compounds, we first investigated the distribution of chemical compounds on the wings of *H. cydno* for comparison with published data on *H. melpomene* (Figure 1). The chemical profile of the wings of the two species were quite different, with few shared major compounds. We found that the bouquet of *H. cydno* was simpler than that of *H. melpomene*, with 7 main compounds versus 21 in *H. melpomene*. *Heliconius cydno* also had a less abundant overall bouquet, with an individual total of 1787 ± 776 ng vs. 3174 ± 1040 ng in *H. melpomene* (Table S1). Most of the main compounds (defined as those occurring in at least 90% of individuals) in *H. cydno* were linear alkanes (4 of 7 compounds), while *H. melpomene* has a more diverse array of different compound classes. Of the five major compounds (> 100ng per individual) in *H. melpomene*, only syringaldehyde was found in similar amounts in *H. cydno*; the major compounds octadecanal, 1-octadecanol, (*Z*)-11-icosenal, and (*Z*)-11-icosenol were absent or found in very low amounts in *H. cydno*. Comparison with previously published data for other *Heliconius* species, all of which lack octadecanal in the large amounts seen in *H. melpomene*, demonstrates that this high level of octadecanal is an evolutionarily derived state in *H. melpomene* (Mann et al. 2017; Darragh et al. 2019b). When focusing on the hindwing androconia of *H. cydno*, only two compounds [syringaldehyde (24.7% of the hindwing androconial bouquet) and (*Z*)-11-icosenol (1.7%)] were specific to this region. For details, please see SI text, Figures S1-S3, and Tables S1-S2.

Electroantennographic responses to con- and heterospecific pheromone stimulus sets

We next investigated the electroantennographic (EAG) response of both species to natural con- and heterospecific pheromone bouquets extracted from adult male butterflies. In general, EAG responses were more pronounced in females than in males, and we did not see a pattern of increased response to conspecific pheromone bouquets

over those of heterospecifics. Females of both *H. melpomene* and *H. cydno* responded more strongly (i.e. showed a larger voltage displacement from the antennal baseline) to both natural wing extracts (Mnat and Cnat, respectively) as compared with the control solvent (DCM+IS) (Figure 2, Figure S6; see Table S3 for statistical details). Males of *H. melpomene* also responded to both wing extracts, while no response was seen in male *H. cydno*, likely due to large inter-individual variation in response in this species-sex combination. Females and males of both species showed equivalent responses to *H. melpomene* and *H. cydno* wing extracts.

We then explored antennal responses to synthetic compound blends. These were based on the most abundant compounds from each species (see Methods, Figure S4). We were able to successfully recapitulate the pheromone of *H. melpomene*, but not that of *H. cydno*. Male and female *H. melpomene* responded equally to the natural *H. melpomene* wing extract and its synthetic wing blend (Msyn) in both stimulus sets (Figures 2 and 3). By contrast, both sexes of *H. cydno* evidenced no increased response to the synthetic *H. cydno* wing blend (Csyn) when compared with the hexane solvent. In all cases this response was lower than to natural *H. cydno* wing extract, indicating that we have not successfully identified its active component(s).

Electroantennographic responses to individual pheromone components from both species

Finally, we explored the responses to individual compounds to identify specific biologically active pheromone components. Only octadecanal differed significantly from the controls in any species-sex combination (Figure 3, Table 1, Table S3), and this difference was seen only in females. As our experiments used concentrations approximately 10x those present in nature (approximately equivalent to the concentrated natural extracts, except (*Z*)-13-docosenal which was at 30x based on prior chemical analysis), this is unlikely to be due to differences in compound abundance in our experiments. No other compound was significantly different from hexane. In female *H. melpomene*, response to octadecanal was stronger than response to the *H. melpomene* synthetic mixture, suggesting a slight inhibitory response due to the presence of other synthetic compounds in the mixture, though no single compound produced this inhibition

in isolation. By contrast, male *H. melpomene* responded equally to the conspecific synthetic mixture and octadecanal (as well as the solvent hexane), and both female and male *H. cydno* responded equally to their conspecific synthetic mixture and both its components (syringaldehyde and henicosane), with no evidence for a synergistic mixture effect. Antennal responses to a given stimulus can change over time, and this may reflect biological processes of neuronal adaptation. Female *H. melpomene* adapted equally quickly to octadecanal and the natural and synthetic *H. melpomene* pheromones, while adaptation to other stimuli was equal to the control, further supporting octadecanal's salience as the main pheromone in *H. melpomen*e (see SI text, Figures S7-S8, and Table S4).

Behavioral response to octadecanal supplementation in *H. melpomene*

We next confirmed a behavioral response to the most physiologically active substance, octadecanal, one of the dominant compounds in *H. melpomene*. A total of 29 behavioral experiments were conducted in *H. melpomene*, with mating observed in 18 (62%); one trial was excluded due to wing damage, leaving 17 successful matings. With our small sample size, we found no evidence that females showed a preference for either treatment, mating with the hexane male 11 times (65%) and with the octadecanal male the remaining 6 times (35%) ($p = 0.332$). However, mating latency (time from experiment onset to mating) was significantly longer for the octadecanal matings (average 88.5 minutes) than for the hexane matings (average 43.7 minutes; $t = 2.7848$, $df = 8.2491$, $p = 0.023$). There was no evidence that this mating latency was due to evaporation of the octadecanal treatment, as there was no detectable drop in octadecanal quantity in the hindwing androconia over the duration of the experiment [comparison of 30 minute and 2 hour treatments (Figure S9)], although some octadecanal was lost initially before the experiment began. Furthermore, little octadecanal rubbed off onto the forewing overlap region. Interestingly, although about 5.5x as much octadecanal as normal should have been present on the wings of treated males, only about 2-2.5x was seen after 30 minutes (the time at which the female would be introduced to the two males), suggesting some of the added octadecanal was lost before the start of the behavioral experiments, perhaps due to oxidization or pheromone hydrolysis, as has been shown in some moths (Ferkovich et al. 1982).

Genetic basis of octadecanal production in *H. melpomene*

Analysis of octadecanal production by F_1 males showed that the *H. cydno* octadecanal phenotype (little to no octadecanal) is dominant over the octadecanal-rich *H. melpomene* phenotype, and octadecanal production segregates in backcrosses to *H. melpomene* (Figure S10). Using the variance within the *H. melpomene* backcross individuals, we calculated a Castle-Wright estimator of 0.81 loci, suggesting a potentially monogenic basis for octadecanal production in *H. melpomene*. Quantitative trait locus (QTL) mapping with 89 individuals from 10 families revealed a single significant peak on chromosome 20 (Figure 4a, Figure S11). The chromosome 20 peak remained significant when kinship was taken into account and explained 41.31% of the difference between the two parent species. Bayesian confidence intervals for the peak on chromosome 20 were identical with and without kinship, spanning a range of 46.9-56.37cM with the peak at 47.66cM (with or without kinship), corresponding to a physical range of 3.4Mb along chromosome 20. To ensure that our findings were replicable across individual families, we also constructed effect plots at the kinship peak for each family separately, and all showed the same directionality (Figure 4b). The peak on chromosome 20 does not overlap with any of the major wing color loci (Jiggins 2017; Van Belleghem et al. 2017), nor does it overlap with mate choice QTL (Merrill et al. 2019). The confidence interval region contains 160 genes, all of which represent potential candidates for octadecanal production. Although octadecanal production appeared recessive in F_1 individuals, there was also some segregation seen in backcrosses to *H. cydno*, so we performed QTL mapping on these additional individuals (127 individuals, 15 families). This analysis recapitulated the peak on chromosome 20 with a similar confidence interval (42.35-54.85cM with a peak at 45.76cM), providing independent support for this QTL peak from an entirely different set of hybrid individuals.

We next evaluated the evidence in support of the 160 genes in this interval to identify top candidates for octadecanal production. In total, 14 were putative fatty-acyl CoA reductases (FARs), which catalyze the conversion of fatty acids to alcohols via a bound aldehyde intermediate (Table S5). Octadecanal is most likely produced via this pathway (Figure 4c), either as a direct product of a FAR-catalyzed conversion of 18-carbon stearic acid (by releasing the bound intermediate directly) or as a product of a further

dehydrogenation of the alcohol intermediate (octadecanol) to the aldehyde product. Two candidate alcohol dehydrogenases, which might catalyze this reaction, were also contained within the region, yielding a total of 16 candidates. To ascertain whether octadecanol might serve as the precursor to octadecanal in *H. melpomene*, we also searched for QTL underlying octadecanol production (Castle-Wright estimator of 0.71 loci), and found a very similar pattern to octadecanal, with a single QTL peak on chromosome 20 (Figure 4d) explaining 25.36% of the difference between the two parent species. This peak broadly overlapped the octadecanal peak, with a much broader confidence interval from 10.91-56.37cm (12.9Mb) and a peak at 51.82cM regardless of whether kinship was taken into account (Figure S11). The 14 FARs in the region are highly clustered, with a set of eight found within a 133kb region. Sequence comparison between the *H. melpomene* and *H. cydno* alleles showed that nearly all of these genes harbor nonsynonymous SNPs between the two species (Table S5, Supplementary Data 1). One gene showed no coding SNPs between the two species; nine had between three and 24 nonsynonymous SNPs; and four had more substantial changes, including missing exons and frameshifts. The final two genes (one alcohol dehydrogenase and one FAR) could not be found in the *H. cydno* genome, and may instead represent annotation or assembly errors in *H. melpomene* or, alternately, deletions in *H. cydno*. Nearly all of the intact genes displayed purifying selection (ω between 0.001 and 0.2459), with only the remaining alcohol dehydrogenase (ω = 1.2478) under positive selection (Table S5). Taken together, these results suggest that either of the alcohol dehydrogenase candidates may underlie the production of bioactive octadecanal from octadecanol in *H. melpomene*, although functional experiments are required to confirm this hypothesis. Alternately, as QTL for octadecanal and its likely precursor octadecanol overlap, a single FAR may be responsible for producing both volatiles.

Discussion

Previous work has shown that male *Heliconius* butterflies use aphrodisiac pheromones during courtship, and the presence of these pheromones is necessary for successful mating to take place (Darragh et al. 2017). However, the identity of the bioactive pheromone components and the genetic basis underlying their production was unknown. Here we demonstrate that two closely related species with strong

reproductive isolation, *H. melpomene* and *H. cydno*, show major differences in chemical bouquets (see also Mann et al. 2017; Darragh et al. 2019b). Strong divergence between closely related species is unusual in Lepidoptera. Instead, pheromone types are typically shared between closely related species, with only subtle differences in similar compounds or differences in ratios of the same compound (Löfstedt et al 2016). Somewhat surprisingly, despite these major differences in putative pheromone signals, we detected no difference in the strength of antennal response to wing chemical bouquets. Nonetheless, we have identified a single compound, octadecanal, which elicited a significant response in females of both species. Somewhat surprisingly, octadecanal was also physiologically active in *H. cydno* females, while it is largely absent from the male *H. cydno* wing bouquet.

These data on identical antennal responses may suggest that the peripheral nervous system of *H. melpomene* and *H. cydno* have not diverged in concert with their male wing chemistry. This is in contrast to similar published studies of other insect species. For example, the moth *Ostrinia nubialis*, whose *E*- and *Z*-strains diverged approximately 75,000-150,000 years ago (Malausa et al. 2007), strains have opposite topologies in the antennal lobe and antennal sensillae (Kárpáti et al. 2007; Koutroumpa et al. 2014). Similar divergence in peripheral nervous system architecture has been seen in *Rhagoletis pomonella* (Frey and Bush 1990; Tait et al. 2016) and *Drosophila mojavensis* (Date et al. 2013; Crowley-Gall et al. 2016) despite much shorter divergence times than between *H. melpomene* and *H. cydno* (approximately 2.1 million years ago, Arias et al 2014; Kozak et al. 2015).

The sensory periphery is only the first of many mechanisms that may influence mate choice, and it is increasingly clear that differences within the brain can be important in mate choice, even when detection mechanisms of the sensory periphery are conserved (Hoke et al. 2008; Hoke et al. 2010; Seeholzer et al. 2018). Our results in *Heliconius* are similar to those observed in *Colias* butterflies. The sister species *C.eurytheme* and *C. philodice* show very similar female electrophysiological responses to the con- and heterospecific pheromone compounds, despite a behavioral effect of treating males with heterospecific pheromones (Grula et al. 1980). *Colias* butterflies, like *Heliconius*, use multiple signals when choosing between mates (Papke et al. 2007), so rapid divergence in

peripheral nervous system elements may not play a role in the evolution of reproductive isolation. In *Heliconius*, where EAG responses are very similar between species, differences in the antennal lobe and higher brain regions (e.g. the mushroom body or lateral protocerebrum, see e.g. Montgomery & Merrill 2017) may account for interspecies differences in mate choice behavior. (). Female *Heliconius* butterflies likely integrate multiple signals (including pheromones, male courtship flights, and visual cues) in these higher brain regions when making the decision to mate.

We have identified octadecanal as the major pheromone component in *H. melpomene* and showed that responses to it are conserved across both species despite its general absence in *H. cydno*. As a fully saturated unbranched compound, octadecanal is unusual in being unrelated to known female pheromone types in moths, which largely use unsaturated or methylated hydrocarbons with or without terminal functional groups (Löfstedt et al. 2016). The activity of octadecanal as a pheromone, however, has been tested behaviorally or electrophysiologically in eight species of Lepidoptera across a variety of families (Tatsuki et al. 1983; Cork et al. 1988; Tumlinson et al. 1989; Ho et al. 1996; McElfresh et al. 2000; Yildizhan et al. 2009; El-Sayed et al. 2011; Pires et al. 2015; Chen et al. 2018). Only in *Cerconota anonella* (Pires et al. 2015), and now in *Heliconius melpomene*, has some electrophysiological and behavioral activity been seen. The closely related hexadecanal is also a major pheromone component in the butterfly *Bicyclus anynana* (Nieberding et al. 2008), and differs from octadecanal only in its origin from palmitic rather than stearic acid and carbon number, so the role of octadecanal is not entirely unexpected.

Male pheromones categorized in other butterflies represent a wide range of chemical classes, including terpenoids (Meinwald et al. 1969; Pliske & Eisner 1969; Andersson et al. 2007), pyrrolizidine alkaloid derivatives (Meinwald et al. 1969; Pliske & Eisner 1969; Nishida et al. 1996; Schulz & Nishida 1996), macrolides (Yildizhan et al. 2009), aromatics (Andersson et al. 2003), fatty acid esters (Grula et al. 1980), and (in *Bicyclus anynana*) unsaturated fatty acid derived compounds more typical of moths (Nieberding et al. 2008). Male moth pheromones follow this wide distribution of chemical classes, with only a few species using fatty acid derived compounds, most notably *Heliothis virescens* which uses octadecanol, hexadecanol, and related compounds but not

the respective aldehydes such as octadecanal (Conner & Iyengar 2016). In contrast to *H. melpomene*, we have failed to discover any physiologically active pheromones in *H. cydno*, perhaps because a minor component or components not tested here is biologically active (McCormick et al. 2014; Chen et al. 2018). Attempts to identify the *H. cydno* pheromone using GC-coupled electroantennographic detection were unfortunately unsuccessful due to technical issues with the setup, and thus the *H. cydno* pheromone remains undescribed.

Intriguingly, despite the strong EAG response, there was a marked negative behavioral response to increased octadecanal in *H. melpomene*. A plausible explanation for the negative behavioral response to octadecanal supplementation is that disruption of the normal mixture ratios of *H. melpomene* may inhibit the female response, as seen in the butterfly *Pieris napi*, where synergistic processing of two volatile components in the male bouquet is necessary for acceptance behavior (Larsdotter-Mellstrom et al. 2016). Octadecanal may also experience a dose-response curve with an aversive response to higher concentrations and an attractive response at lower ones. Potential mixture or dosage effects suggest that female *H. melpomene* may use octadecanal quantity or relative abundance to assess male quality or choose between courting males. The increased mating latency with octadecanal-treated males may reflect females undergoing a period of adjustment, either in the peripheral or central nervous system, to the higher dose of octadecanal; this would be consistent with our results showing long-term adaptation to octadecanal. We remain uncertain of what effect, if any, octadecanal would have on the behavior of *H. cydno*, where it is largely absent from the male pheromone bouquet, as we were unable to rear an adequate number of *H. cydno* for behavioral trials. It may be used to avoid courtship with *H. melpomene*, supporting other divergent signals such as color pattern in maintaining reproductive isolation between the two species (Jiggins et al. 2001).

Given the strong physiological and behavioral response to octadecanal, and its possible role in reproductive isolation between *H. melpomene* and *H. cydno*, we studied the genetic basis of differences between the two species. Fatty-acid derived compounds comprise the largest category of Lepidoptera sex pheromones (Ando and Yamakawa 2011), and are produced from fatty acyl-CoA precursors via the action of several

enzymes. Since these pheromones are secondary metabolites derived from primary metabolic pathways, their production is likely to be relatively labile in evolutionary terms, allowing simple genetic changes to drive the wide diversity of lepidopteran sex pheromones. Even though we have a broad knowledge of pheromone diversity in Lepidoptera, our understanding of the genetics of pheromone biosynthesis is relatively weak. Pheromone gland-specific fatty acyl-CoA reductases have been identified in a number of moth species, although most are identified solely on transcriptomic analysis of the gland without functional characterization (Groot et al 2016, Löfstedt et al 2016). In the moth *Ostrinia* and butterfly *Bicyclus*, FARs involved in male pheromone biosynthesis have been identified and shown to use the same biosynthetic pathway as female pheromones (Lassance & Löfstedt 2009, Lienard et al 2014). Pheromone-producing alcohol oxidases, which potentially catalyze the conversion of antennally inactive octadecanol to the active component octadecanal, have not yet been described in any insect to our knowledge.

Using a QTL mapping approach, we have shown that the production of octadecanal has a relatively simple genetic basis, with a region on chromosome 20 corresponding to production of both octadecanal and its likely precursor octadecanol in *Heliconius*. This locus therefore likely represents a region under divergent selection between *H. melpomene* and *H. cydno* that is unlinked to previously identified species differences in color and mate choice (Jiggins 2017; Merrill et al. 2019). Patterns of F_{ST} between the species are highly heterogeneous and were not especially informative in further delimiting the locus (data from Martin et al. 2013). Due to our small mapping population, the confidence intervals for these QTL therefore remain large: the octadecanal QTL spans 3.4Mb and contains 160 genes. Of these, we identified 16 likely candidate genes based on known biosynthetic pathways in moths and the butterfly *Bicyclus anynana* (Liénard et al. 2014): 14 fatty acyl-CoA reductases and two alcohol dehydrogenases. Fatty acyl-CoA reductases have previously been identified in *H. melpomene* by (Liénard et al. 2014), who noted lineage-specific duplications within *H. melpomene* on two scaffolds corresponding to *H. melpomene* chromosomes 19 and 20. All but one of the candidate FARs found on chromosome 20 were identified by Liénard et al., but all fall outside their clade of pheromone gland FARs. The identified *Bicyclus* FAR

that produces hexadecanol does not also produce the major pheromone hexadecanal, implying the presence of an additional as yet undescribed alcohol dehydrogenase in *Bicyclus*, and the biochemical similarity between hexadecanal and octadecanal suggests *Heliconius* may also use an alcohol dehydrogenase to produce octadecanal. By contrast, the overlapping octadecanol and octadecanal QTL on chromosome 20 in *Heliconius* suggest the presence of a bifunctional FAR that produces both the alcohol and aldehyde together, or alternately tight linkage of separate FAR and alcohol dehydrogenase genes. Further studies, including functional assays and location of wing pheromone biosynthesis, will be required to tease apart our potential candidates.

The presence of a single large-effect QTL for octadecanal production is not surprising, as large-effect loci have been seen in pheromone production in various moth species (Groot et al. 2016, Haynes et al. 2016). What is more surprising is that species differences in moths largely are the result of minor variations in similar compounds or compound ratios, while the production of both octadecanal and its precursor octadecanol is essentially absent in *H. cydno*. Nevertheless, despite the recruitment of stearic acid into this novel product, we see only a single QTL, potentially due to a single gene or tight linkage of two or more biosynthetic genes. The octadecanal locus on chromosome 20 does not overlap with any of the known genes involved in color pattern and mate choice in *Heliconius*, which all lie on other chromosomes (Jiggins 2017), and notably there is no overlap with the *optix* color pattern gene or previously described mate choice QTL to which it is tightly linked (Merrill et al. 2019). Tight linkage of loci for traits under divergent selection and those contributing to premating isolation should facilitate speciation (Felsenstein 1981; Merrill et al. 2010; Smadja and Butlin 2011) but based on our data there is no linkage between olfactory cues and divergent warning patterns in *Heliconius melpomene* and *H. cydno*. It is possible that olfaction does not play a significant role in reproductive isolation. Nonetheless, other color pattern loci are scattered across the *Heliconius* genome, rather than being tightly linked. Instead of acting as a color pattern, mate choice, and pheromone supergene, the loci responsible for these traits are mostly unlinked. Perhaps the selection favoring genetic linkage is weak in these species now that speciation is nearly complete (see also Davey et al. 2017).

Our studies of the electrophysiological and behavioral responses of *Heliconius* butterflies and the genetic basis of pheromone production add to the growing body of literature suggesting that pheromonal communication in Lepidoptera is not limited to nocturnal moths but can be found in day-flying butterflies that also use striking visual signals. *Heliconius* butterflies can detect con- and heterospecific wing compound bouquets, and a major component, octadecanal, is physiologically and behaviorally active in *H. melpomene* and its genetic basis appears relatively simple, consistent with other pheromone shifts found in insects (Symonds and Elgar 2007; Smadja & Butlin 2009). Along with their striking wing color patterns, male *Heliconius* use chemistry to influence female mate choice, combining courtship behaviors, and chemistry in a dance to elicit female mating responses (Klein and de Araújo 2010; Mérot et al. 2015). Despite our human bias towards visual signals, we are now beginning to understand how such visually striking butterflies communicate using chemistry.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Andersson, J., Borg-Karlson, A.-K., Vongvanich, N., and C. Wiklund. 2007. Male sex pheromone release and female mate choice in a butterfly. J. Exp. Biol. 210:964-970.

Andersson, J., Borg-Karlson, A.-K., and C. Wiklund. 2003. Antiaphrodisiacs in pierid butterflies: A theme with variation! J.Chem. Ecol. 29:1489–1499.

Ando, T., and R. Yamakawa. 2011. Analyses of lepidopteran sex pheromones by mass spectrometry. Trends Anal. Chem. 30:990-1002.

Arias, C. F., Salazar, C., Rosales, C., Kronforst, M. R., Linares, M., Bermingham, E., and W. O. McMillan. 2014. Phylogeography of *Heliconius cydno* and its closest relatives: disentangling their origin and diversification. Molecular Ecology 23:4137-4152.

Broman, K. W., Gatti, D. M., Simecek, P., Furlotte, N. A., Prins, P., Sen, Ś., Yandell, B. S., and G. A. Churchill. 2018. R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multi-parent populations. Genetics 211:495-502.

Cha, J. S., and H. C. Brown. 1993. Reaction of sodium aluminum hydride with selected organic compounds containing representative functional groups. Comparison of the reducing characteristics of lithium and sodium aluminum hydrides. J. Org. Chem. 58:4727–4731.

Challis, R. J., Kumar, S., Dasmahapatra, K. K., Jiggins, C. D., and M. Blaxter. 2016. Lepbase: the Lepidopteran genome database. bioRxiv 056994v1.

Chen, L., and H. Y. Fadamiro. 2007. Behavioral and electroantennogram responses of phorid fly *Pseudacteon tricuspis* (Diptera: Phoridae) to Red Imported Fire Ant *Solenopsis invicta* odor and trail pheromone. J. Insect Behav. 20:267-287.

Chen, Q.-H., Zhu F., Tian, Z., Zhang, W.-M., Guo, R., Liu, W., Pan, L., and Y. Du. 2018. Minor components play an important role in interspecific recognition of insects: a basis to pheromone based electronic monitoring tools for rice pests. Insects 9:192.

Connor, W. E., and V. K. Iyengar. 2016. Male pheromones in moths: reproductive isolation, sexy sons, and good genes. In *Pheromone Communication in Moths*, J. D. Allison and R. Cardé, ed. University of California Press.

Cork, A., Chamberlain, D. J., Beevor, P. S., Hall, D. R., Nesbitt, B. F., Campion, D. G., and M. R. Attique. 1988. Components of female sex pheromone of spotted bollworm, *Earias vittella* F. (Lepidoptera: Noctuidae): identification and field evaluation in Pakistan. J. Chem. Ecol. 14:929-945.

Crowley-Gall, A., Date, P., Han, C., Rhodes, N., Andolfatto, P., Layne, J. E., and S. M. Rollmann. 2016. Population differences in olfaction accompany host shift in *Drosophila mojavensis*. Proc. R. Soc. B 283:20161562.

Darragh, K., Vanjari, S., Mann, F., Gonzalez-Rojas, M. F., Morrison, C. R., Salazar, C., Pardo-Diaz, C., Merrill, R. M., McMillan, W. O., Schulz, S., and C. D. Jiggins. 2017. Male sex pheromone components in *Heliconius* butterflies released by the androconia affect female choice. PeerJ 5:e3953.

Darragh, K., Byers, K. J. R. P., Merrill, R. M., McMillan, W. O., Schulz, S., and C. D. Jiggins. 2019a. Male pheromone composition depends on larval but not adult diet in *Heliconius melpomene*. Ecol. Entomol. 44:397-405.

Darragh, K., Montejo-Kovacevich G., Kozak, K. M., Morrison, C. R., Figueiredo, C. M. E., Ready, J. S., Salazar, C., Linares, M., Byers, K. J. R. P., Merrill, R. M., McMillan, W. O., Schulz, S., and C. D. Jiggins. 2019b. Species specificity and intraspecific variation in the chemical profiles of *Heliconius* butterflies across a large geographic range. bioRxiv 573469v2.

Date, P., Dweck, H. K. M., Stensmyr, M. C., Shann, J., Hansson, B. S., and S. M. Rollmann. 2013. Divergence in olfactory host plant preference in *D. mojavensis* in response to cactus host use. PLOS ONE 8:e70027.

Davey, J. W., Barker S. L., Rastas, P. M., Pinharanda, A., Martin, S. H., Durbin, R., McMillan, W. O., Merrill, R. M., and C. D. Jiggins. 2017. No evidence for maintenance of a sympatric *Heliconius* species barrier by chromosomal inversions. Evol. Lett. 1:138-154.

Dopman, E. B., Robbins, P. S., and A. Seaman. 2010. Components of reproductive isolation between North American pheromone strains of the European Corn Borer. Evolution 64:881-902.

El-Sayed, A. M., Mitchell, V. J., Manning, L. M., and D. M. Suckling. 2011. New sex pheromone blend for the Lightbrown Apple Moth, *Epiphyas postvittana*. J. Chem. Ecol. 37:640-646.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution 35:124-138.

Ferkovich, S. M., Oliver, J. E., and C. Dillard. 1982. Pheromone hydrolysis by cuticular and interior esterases of the antennae, legs, and wings of the cabbage looper moth, *Trichoplusia ni* (Hübner). J. Chem. Ecol. 8:859-866.

Fox, J., and S. Weisberg. 2011. An R companion to applied regression, second edition. Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Frey, J. E., and G. L. Bush. 1990. *Rhagoletis* sibling species and host races differ in host odor recognition. Entomol. Exp. Appl. 57:123-131.

Groot, A. T., Dekker, T., and D. G. Heckel (2016) The genetic basis of pheromone evolution in moths. Annu. Rev. Entomol. 61:99-117.

Grula, J. W., McChesney, J. D., and O. R. Taylor, Jr. 1980. Aphrodisiac pheromones of the sulfur butterflies *Colias eurytheme* and *C. philodice* (Lepidoptera, Pieridae). J. Chem. Ecol. 6:241-256.

Ho, H. Y., Tao, Y. T., Tsai, R. S., Wu, Y. L., Tseng, H. K., and Y. S. Chow. 1996. Isolation, identification, and synthesis of sex pheromone components of female tea cluster caterpillar, *Andraca bipunctata* Walker (Lepidoptera: Bombycidae) in Taiwan. J. Chem. Ecol. 22:271-285.

Hoke, K. L., Ryan, M. J., and W. Wilczynski. 2008. Candidate neural locus for sex differences in reproductive decisions. Biol. Lett. 4**,** 518–521.

Hoke, K. L., Ryan, M. J., and W. Wilczynski. 2010. Sexually dimorphic sensory gating drives behavioral differences in tungara frogs. J. Exp. Biol. 213**,** 3463–3472.

Jiggins, C. D. 2017. The ecology and evolution of *Heliconius* butterflies. Oxford University Press.

Jiggins, C. D., Naisbit, R. E., Coe, R. L., and J. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411:302-305.

Jones, C. D. 2001. Extension of the Castle–Wright effective factor estimator to sex linkage and haplodiploidy. J. Hered. 92:274-276.

Kárpáti, Z., Dekker, T., and B. S. Hansson. 2007. Reversed functional topology in the antennal lobe of the male European corn borer. J. Exp. Biol. 211:2841-2848.

Klein, A. L., and A. M. de Araújo. 2010. Courtship behavior of *Heliconius erato phyllis* (Lepidoptera, Nymphalidae) towards virgin and mated females: conflict between attraction and repulsion signals? J. Ethology 28:409-420.

Kochansky, J., Cardé, R. T., Liebherr, J., and W. L. Roelofs. 1975. Sex pheromone of the European corn borer, *Ostrinia nubialis* (Lepidoptera: Pyralidae), in New York. J. Chem. Ecol. 1:225-231.

Koller, M. 2016. robustlmm: An R package for robust estimation of linear mixed-effects models. J. Stat. Software 75:1-24.

Koutroumpa, F. A., Kárpáti, Z., Monsempes, C., Hill, S. R., Hansson, B. S., Jacquin-Joly, E., Krieger, J., and T. Dekker. 2014. Shifts in sensory neuron identity parallel differences in pheromone preference in the European corn borer. Frontiers Ecol. Evol. 2:65.

Kozak, K. M., Wahlberg, N., Neild, A. F. E., Dasmahapatra, K. K., Mallet, J., and C. D. Jiggins. 2015. Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. Syst. Biol. 64:505-524.

Larsdotter-Mellstrom, H., Eriksson, K., Liblikas I, I., Wiklund, C., Borg-Karlson, A. K., Nylin, S., Janz, N., and M. A. Carlsson. 2016. It's all in the mix: blend-specific behavioral response to a sexual pheromone in a butterfly. Frontiers Physiol. 7:68.

Lassance, J.-M., Groot, A. T., Liénard, M. A., Antony, B., Borgwardt, C., Andersson, F., Hedenström, E., Heckel, D. G., and C. Löfstedt. 2010. Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. Nature 466:486-489.

Lassance, J.-M., and C. Löfstedt. 2009. Concerted evolution of male and female display traits in the European corn borer, *Ostrinia nubialis*. *BMC Biol.* 7:10.

Lenth, R. 2018. emmeans: estimated marginal means, aka least-squares means. R package version 1.3.1. URL https://CRAN.R-project.org/package=emmeans

Li, H., and R. Durbin. 2011. Inference of human population history from individual whole-genome sequences. Nature 475:493-496.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arxiv 1303.3997.

Liénard, M. A., Wang, H.-L., Lassance, J.-M., and C. Löfstedt. 2014. Sex pheromone biosynthetic pathways are conserved between moths and the butterfly *Bicyclus anynana*. Nat. Comms. 5:3957.

Löfstedt, C., Wahlberg, N., and J. G. Millar. 2016. Evolutionary patterns of pheromone diversity in Lepidoptera. In *Pheromone Communication in Moths*, J. D. Allison and R. Cardé, ed. University of California Press.

Malausa, T., Leniaud, L., Martin, J.-F., Audiot, P., Bourguet, D., Ponsard, S., Lee, S.-F., Harrison, R. G., and E. Dopman. 2007. Molecular differentiation at nuclear loci in French host races of the European corn borer (*Ostrinia nubilalis*). Genetics 176:2343-2355.

Mann, F., Vanjari, S., Rosser, N., Mann, S., Dasmahapatra, K. K., Corbin, C., Linares, M., Pardo-Diaz, C., Salazar, C., Jiggins, C. D., and S. Schulz. 2017. The scent chemistry of *Heliconius* wing androconia. J. Chem. Ecol. 43:843-857.

Martin, S. H., Dasmahapatra K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., and C. D. Jiggins. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 23:1817-1828.

McCormick, A. C., Gershenzon, J., and S. B. Unsicker. 2014. Little peaks with big effects: establishing the role of minor plant volatiles in plant–insect interactions. Plant Cell Env. 37:1836-1844.

McElfresh, J. S., Chen, X., Ross, D. W., and J. G. Millar. 2000. Sex pheromone blend of the pandora moth (Lepidoptera: Saturniidae), an outbreak pest in pine forests (Pinaceae). Canadian Entomologist 132:775-787.

Meinwald, J., Meinwald, Y. C., and P. H. Mazzocchi. 1969. Sex pheromone of the queen butterfly: chemistry. Science 164:1174-1175.

Merrill, R. M., Rastas P., Martin S. H., Melo, M. C., Barker, S., Davey, J., McMillan, W. O., and C. D. Jiggins. 2019. Genetic dissection of assortative mating behavior. PLOS Biology 17:e2005902.

Merrill, R. M., Van Schooten, B., Scott, J. A., and C. D. Jiggins. 2010. Pervasive genetic associations between traits causing reproductive isolation in *Heliconius* butterflies. Proc. R. Soc. Lond. B 278:511-518.

Montgomery, S. H., and R. M. Merrill. 2017. Divergence in brain composition during the early stages of ecological specialization in *Heliconius* butterflies. J. Evol. Biol. 30:571-582.

More, J. D., and N. S. Finney. 2002. A simple and advantageous protocol for the oxidation of alcohols with o-iodoxybenzoic acid (IBX). Org. Lett. 4:3001–3003.

Mérot, C., Frérot, B., Leppik, E., and M. Joron. 2015. Beyond magic traits: Multimodal mating cues in *Heliconius* butterflies. Evolution 69:2891-2904.

Naisbit, R. E., Jiggins, C. D., and J. Mallet. 2003. Mimicry: developmental genes that contribute to speciation. Evol. and Devel. 5:269-280.

Nieberding, C. M., de Vos, H., Schneider, M. V., Lassance, J.-M., Estramil, N., Andersson, J., Bång, J., Hedenström, E., Löfstedt, C., and P. M. Brakefield. 2008. The male sex pheromone of the butterfly *Bicyclus anynana*: towards an evolutionary analysis. PLOS ONE 3:e2751.

Nishida, R., Schulz, S., Kim, C. S., Fukami, H., Kuwahara, Y., Honda, K., and N. Hayashi. 1996. Male sex pheromone of a giant danaine butterfly, *Idea leuconoe*. J. Chem. Ecol. 22:949-972.

Papke, R. S., Kemp, D. J., and R. L. Rutowski. 2007. Multimodal signaling: structural ultraviolet reflectance predicts male mating success better than pheromones in the butterfly *Colias eurytheme*. Anim. Behav. 73:47-54.

Peakall, R., Ebert, D., Poldy, J., Barrow, R. A., Francke, W., Bower, C. C., and F. P. Schiestl. 2010. Pollinator specificity, floral odour chemistry and the phylogeny of Australian sexually deceptive *Chiloglottis* orchids: implications for pollinator-driven speciation. New Phytol. 188:437-450.

Pessoa Pinharanda, A. L. 2017. The genomic basis of species barriers in *Heliconius* butterflies (Doctoral thesis). https://doi.org/10.17863/CAM.16857

Picelli S., Björklund, Å. K., Reinius, B., Sagasser, S., Winberg, G., and R. Sandberg. 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 24:2033-2040.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team. 2018. nlme: linear and nonlinear mixed effects models. R package version 3.1-137, URL: https://CRAN.R-project.org/package=nlme

Pires, E. V., Mendonça, A. de L., Vaníčková, L., Serra, N. S. J., da Silva, R. de C. C., dos Santos, D. C., Campos, R. da S., Sant'Ana, A. E. G., and R. R. do Nascimento. 2015. Identification and field and laboratory tests of the sex pheromone of *Cerconota anonella* Sepp. (Lepidoptera: Oecophoridae). J. Appl. Entomol. 140:72-80.

Plenderleith, M., van Oosterhout, C., Robinson, R. L., and G. F. Turner. 2005. Female preference for conspecific males based on olfactory cues in a Lake Malawi cichlid fish. Biol. Lett. 1:411-414.

Pliske, T. E., and T. Eisner. 1969. Sex pheromone of the queen butterfly: biology. Science 164:1170-1172.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rastas, P. 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics 33:3726-3732.

Schulz, S., and R. Nishida. 1996. The pheromone system of the male danaine butterfly, *Idea leuconoe*. Bioorg. Med. Chem. 4:341-349.

Seeholzer, L. F., Seppo, M., Stern, D. L., and V. Ruta. 2018. Evolution of a central neural circuit underlies *Drosophila* mate preferences. Nature 559:564-569.

Shahandeh, M. P., Pischedda, A., and T. L. Turner. 2017. Male mate choice via cuticular hydrocarbon pheromones drives reproductive isolation between *Drosophila* species. Evolution 72:123-135.

Smadja, C. M., and R. K. Butlin. 2009. On the scent of speciation: the chemosensory system and its role in premating isolation. Heredity 102:77-97.

Smadja, C. M., and R. K. Butlin. 2011. A framework for comparing processes of speciation in the presence of gene flow. Mol. Ecol. 20:5123-5140.

Stengl, M. 2010. Pheromone transduction in moths. Frontiers Cell. Neurosci. 4:133.

Symonds, M. R. E., and M. A. Elgar. 2007. The evolution of pheromone diversity. Trends Ecol. Evol. 23:220-228.

Tait, C., Batra, S., Ramaswamy, S. S., Feder, J. L., and S. B. Olsson. 2016. Sensory specificity and speciation: a potential neuronal pathway for host fruit odour discrimination in *Rhagoletis pomonella*. Proc. R. Soc. Lond. B 283:20162101.

Tatsuki, S., Kurihara, M., Usui, K., Ohguchi, Y., Uchiumi, K., Arai, K., Yabuki, S., and F. Tanaka. 1983. Sex pheromone of the rice stem borer, *Chilo suppressalis* (Walker) (Lepidoptera: Pyralidae): the third component, Z-9-hexadecenal. Appl. Entomol. Zool. 18:443-446.

Tumlinson, J. H., Brennan, M. M., Doolittle, R. E., Mitchell, E. R., Brabham, A., Mazomenos, B. E., Baumhover, A. H., and D. M. Jackson. 1989. Identification of a pheromone blend attractive to *Manduca sexta* (L.) males in a wind tunnel. Arch. Insect Biochem. Physiol. 10:255-271.

Van Belleghem, S. M., Rastas P., Papanicolaou, A., Martin, S. H., Arias, C. F., Supple, M. A., Hanly, J. J., Mallet, J., Lewis, J. J., Hines, H. M., Ruiz, M., Salazar, C., Linares, M., Moreira, G. R. P., Jiggins, C. D., Counterman, B. A., McMillan, W. O., and R. Papa. 2017. Complex modular architecture around a simple toolkit of wing pattern genes. Nature Ecol. Evol. 1:0052.

Vane-Wright, R. I., and M. Boppré. 1993. Visual and chemical signaling in butterflies: functional and phylogenetic perspectives. Phil. Trans. R. Soc. Lond. B 340:197-205.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S, fourth edition. Springer.

Via, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Phil. Trans. R. Soc. Lond. B 367:451-460.

Yang, Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586-1591.

Yew, J. Y., and H. Chung. 2015. Insect pheromones: An overview of function, form, and discovery. Prog. Lipid Res. 59:88-105.

Yildizhan, S., van Loon, J., Sramkova, A., Ayasse M., Arsene, C., ten Broeke, C., and S. Schulz. 2009. Aphrodisiac pheromones from the wings of the Small Cabbage White and Large Cabbage White butterflies, *Pieris rapae* and *Pieris brassicae*. ChemBioChem 10:1666-1677.

Zufall, F., and T. Leinders-Zufall. 2000. The cellular and molecular basis of odor adaptation. Chem. Senses 25:473-481.

Figures

Figure 1: Androconial chemistry of *Heliconius melpomene* and *H. cydno*. a: dorsal forewing and hindwing of each species showing the silvery androconial region of the hindwing used during male courtship. b: Total ion chromatogram of *H. melpomene* (top) and *H. cydno* (bottom) wing androconia. P1: syringaldehyde; P2: octadecanal; P3: 1-octadecanol; P4: henicosane; P5: (*Z*)-11-icosenal; P6: (*Z*)-11-icosenol; P7: (*Z*)-13-docosenal; IS: internal standard (2-tetradecyl acetate); x: contaminant; C21: henicosane; C22: docosane; C23: tricosane.

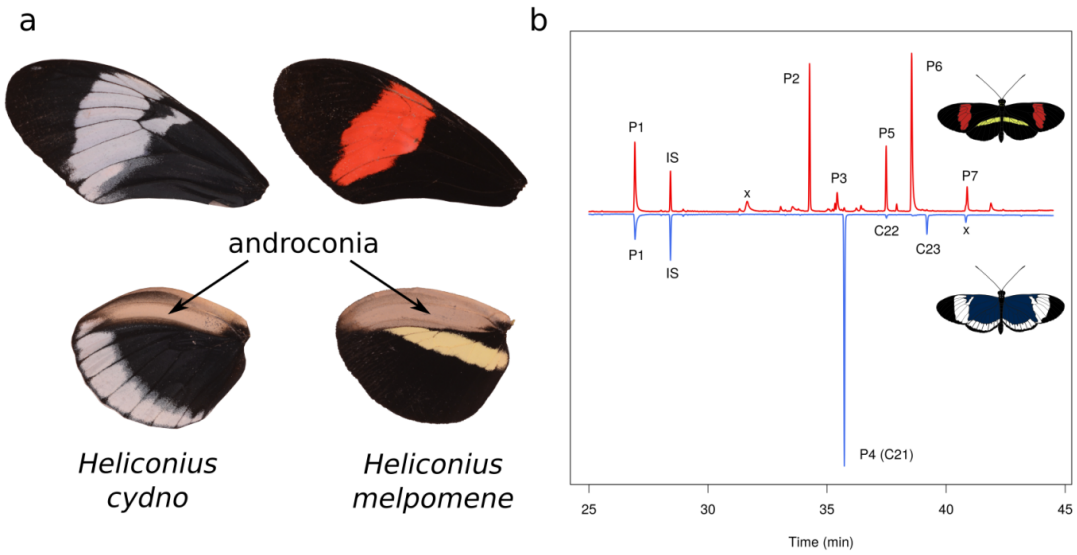


Figure 2: Electroantennographic responses of *Heliconius* butterflies to conspecific and heterospecific wing extracts. Stimuli: air (white), dichloromethane plus 2-tetradecyl acetate (internal standard) (dark gray), hexane (light gray), *Lantana* extract (green), natural male *H. melpomene* wing extract (red), synthetic *H. melpomene* blend (pink), natural male *H. cydno* wing extract (dark blue), synthetic *H. cydno* blend (blue). Bars: average of normalized corrected amplitude ± standard deviation.

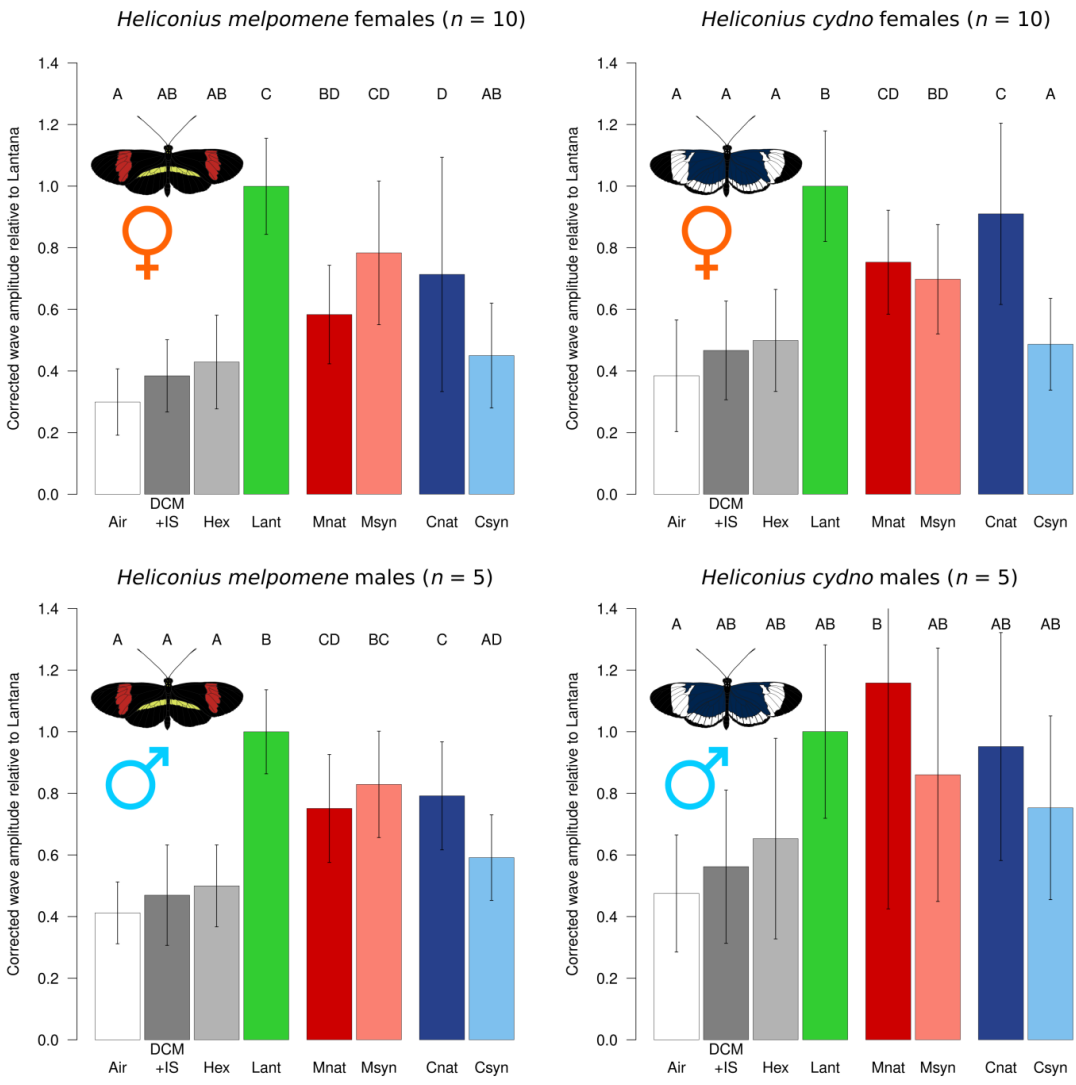


Figure 3: Electroantennographic responses of *Heliconius* butterflies to synthetic wing compounds. Stimuli: air (white), dichloromethane plus 2-tetradecyl acetate (internal standard) (dark gray), hexane (light gray), *Lantana* extract (green), natural male *H. melpomene* or *H. cydno* wing extract (red or dark blue respectively), synthetic *H. melpomene* or *H. cydno* blend (pink or blue respectively). P1 (Syr): syringaldehyde; P2 (18O): octadecanal; P3 (18OH): 1-octadecanol; P4 (C21): henicosane; P5 (20O): (*Z*)-11-icosenal; P6 (20OH): (*Z*)-11-icosenol; P7 (22O): (*Z*)-13-docosenal. Bars: average of normalized corrected amplitude ± standard deviation. Light pink: compound part of synthetic *H. melpomene* blend; light blue: compound part of synthetic *H. cydno* blend; purple: compound in both *H. melpomene* and *H. cydno* synthetic blends.

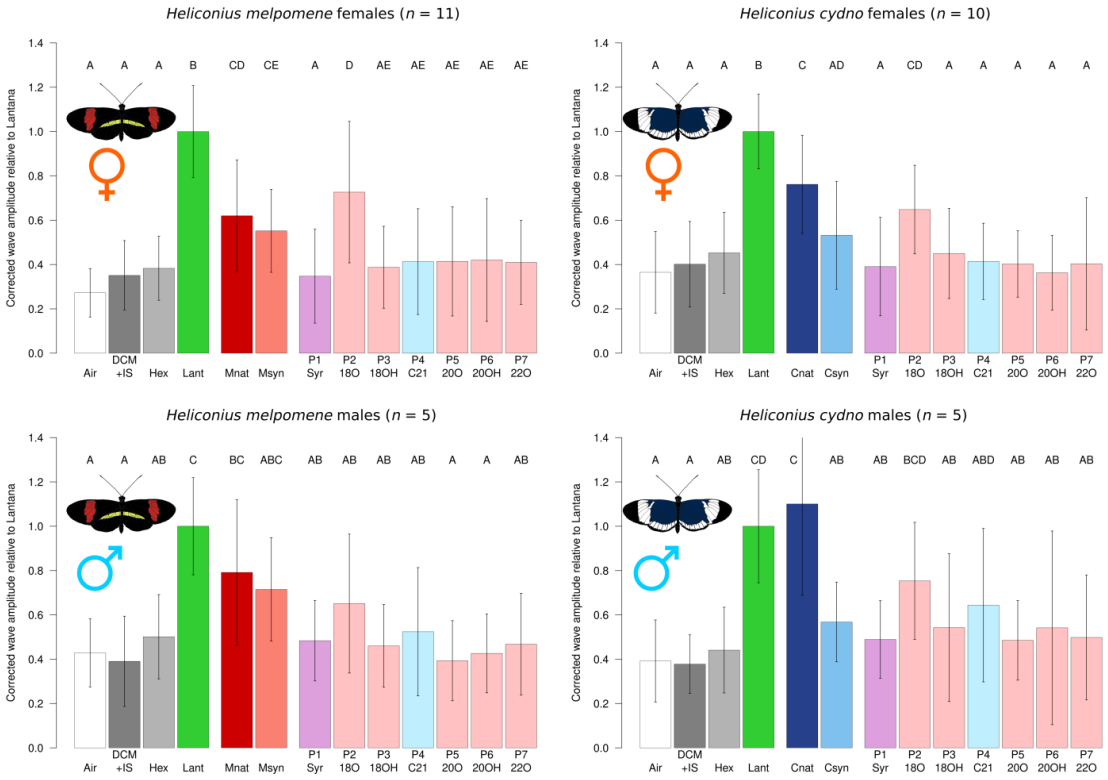
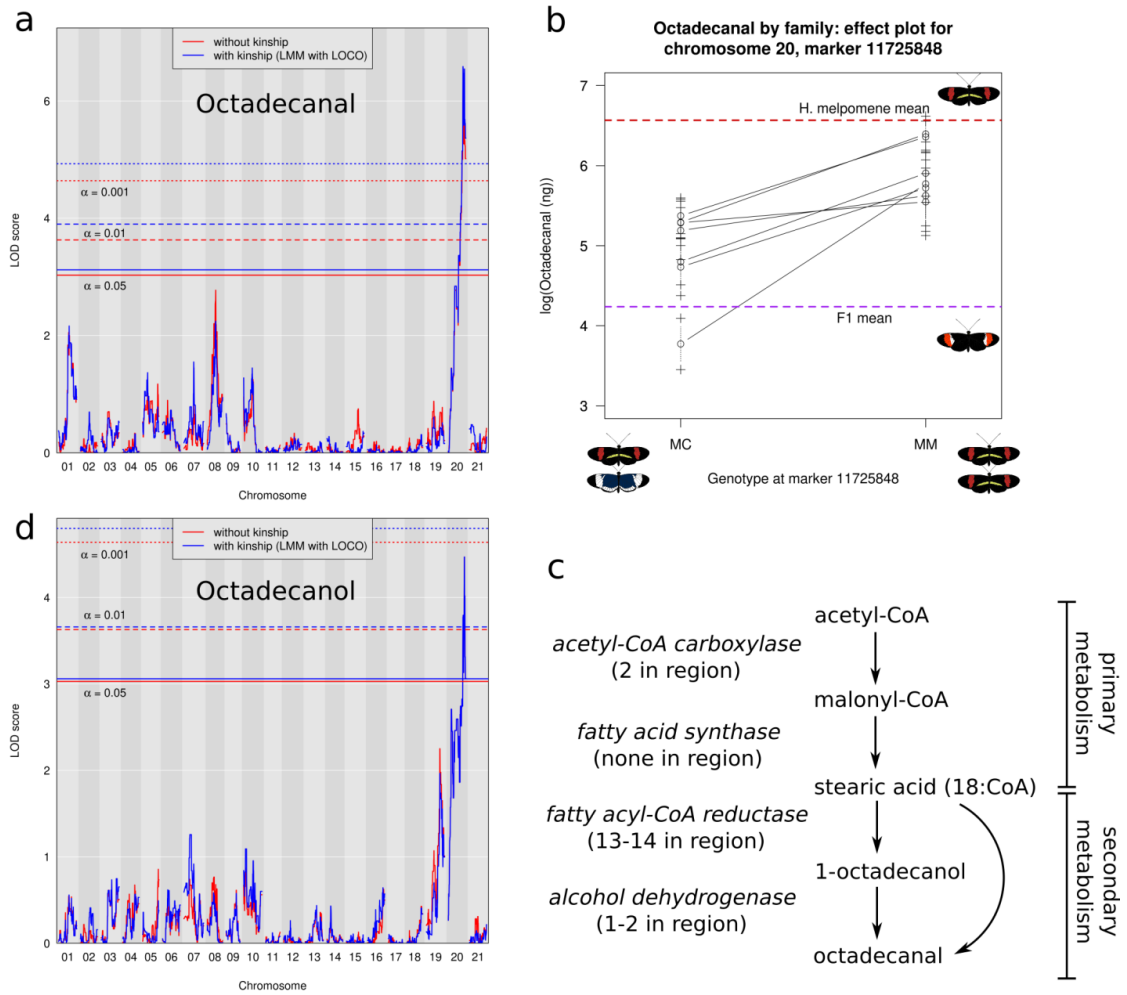


Figure 4: QTL mapping of octadecanal and octadecanol production in *Heliconius melpomene*. a: QTL map for production of octadecanal. b: Effect plots at the peak of the locus on chromosome 20 for the seven individual backcross mapping families with at least 5 individuals. c: Potential biosynthetic pathway for octadecanal production. d: QTL map for production of octadecanol, a potential precursor of octadecanal.



Supplemental Results

Androconial chemical bouquets of *Heliconius melpomene* and *H. cydno*

The analysis of *H. melpomene* included a reanalysis of 18 samples originally published in (Darragh et al. 2017,2019a), as well as 13 previously unpublished samples, for a total of 31. The analysis of *H. cydno* used 26 unpublished samples, eight of which were additionally used for the wing region comparison. Overall, *H. melpomene* produces far more compounds in both amount (3174 ± 1040 ng vs. 1787 ± 776 ng in *H. cydno*, Table S1) and compound diversity (53 compounds found in at least 1/3 of samples vs. 31 compounds in *H. cydno*). Approximately half of all compounds were shared between species (24 of 60 total compounds found in at least 1/3 of samples of both species); of the remainder, the majority of compounds were *H. melpomene*-specific (29 compounds), with a more limited set of *H. cydno*-specific compounds (7 compounds). When grouping compounds by chemical class (Figure S2), *H. melpomene* produced a higher total amount of aromatics and alkenes than *H. cydno*, while *H. cydno* produced more total amount of terpenoids. On a relative basis, the bouquet of *H. melpomene* had a higher percentage of alkenes than *H. cydno*, whereas *H. cydno* was dominated by alkanes. Of the 11 compounds comprising at least 1% of either species' overall bouquet, all but two (the straight-chain alkanes henicosane and tricosane) were found in equal or higher amounts in *H. melpomene* (Figure S1). In *H. cydno*, two compounds (syringaldehyde and henicosane) together made up more than 90% of the overall compound bouquet.

In *H. melpomene*, several compounds are found exclusively or nearly exclusively in the hindwing androconia region (Darragh et al. 2017). We performed a similar assay using eight of the *H. cydno* samples with all four wing regions (hindwing androconia, forewing overlap region, rest of hindwing, rest of forewing) dissected out. Relative to wing region area, the hindwing androconia produced higher amounts of syringaldehyde and (Z)-11-icosenol than the rest of the wing (Figure S3A), while the other five compounds were found in high amounts in any wing region and were relatively equivalent between the androconia and at least two other wing regions. When compound amounts were not corrected for wing region area, the effects were similar (Figure S3B), except for minor differences between other wing regions. Interestingly, there appeared to be a tradeoff

between amounts of syringaldehyde and homovanillyl alcohol across wing regions; as these compounds are structurally similar aromatics this may reflect a shared precursor pool with each wing region being biased towards one or the other product.

Long-term adaptation to pheromone extracts and synthetic compounds

Long-term adaptation (LTA; across the duration of stimulus presentation) was quite common, particularly in responses to natural and synthetic wing pheromone extracts (Figure S7). Adaptation was more common in females, with female *H. melpomene* showing LTA to all stimuli with the exception of henicosane (the only synthetic component tested that is found in low amounts in *H. melpomene*) and disagreement between stimulus sets for air (not shown) and DCM+IS. Female *H. cydno* were similar, with LTA to all stimuli except henicosane, (*Z*)-11-icosenol, and disagreement between stimulus sets for air (not shown). Males showed LTA to fewer stimuli, particularly to synthetic pheromone components; *H. melpomene* males only had LTA to octadecanal and henicosane, and *H. cydno* males only to octadecanal, (*Z*)-11-icosenol, and (*Z*)-13-docosenal. The only natural or synthetic mixture that did not show LTA across all species-sex pairs was *H. melpomene* natural wing (disagreement between stimulus sets in *H. melpomene* males). For synthetic pheromone components, the only disagreements between species within a sex were in (*Z*)-11-icosenol (for females) and henicosane, (*Z*)-11-icosenol, and (*Z*)-13-docosenal (for males).

Long-term adaptation was often stronger to natural stimuli when compared with the two solvents (hexane and DCM+IS). In *H. melpomene* females, adaptation to *H. melpomene* natural and synthetic wing pheromone was stronger than adaptation to DCM+IS, while adaptation to both natural and synthetic *H. cydno* wing pheromone was equivalent to the relevant solvent (DCM+IS and hexane respectively). The same pattern was found in *H. melpomene* males. In *H. cydno* females, only synthetic *H. cydno* wing pheromone showed equivalent LTA to its hexane solvent, while LTA to all other stimuli was stronger than their respective solvents. The pattern differed in *H. cydno* males, with only natural wing extracts from both species having stronger LTA than their solvent (DCM+IS). By contrast with natural stimuli, LTA to pheromone components was often equivalent to the two solvents. In *H. melpomene* females, only octadecanal differed from

the hexane solvent, and the strength of LTA to octadecanal was identical to that of natural and synthetic *H. melpomene* wing pheromone. Other species-sex combinations showed no obvious pattern in comparison to the solvents, although both sexes of *H. cydno* showed stronger LTA to *H. cydno* wing pheromone than to its DCM+IS solvent.

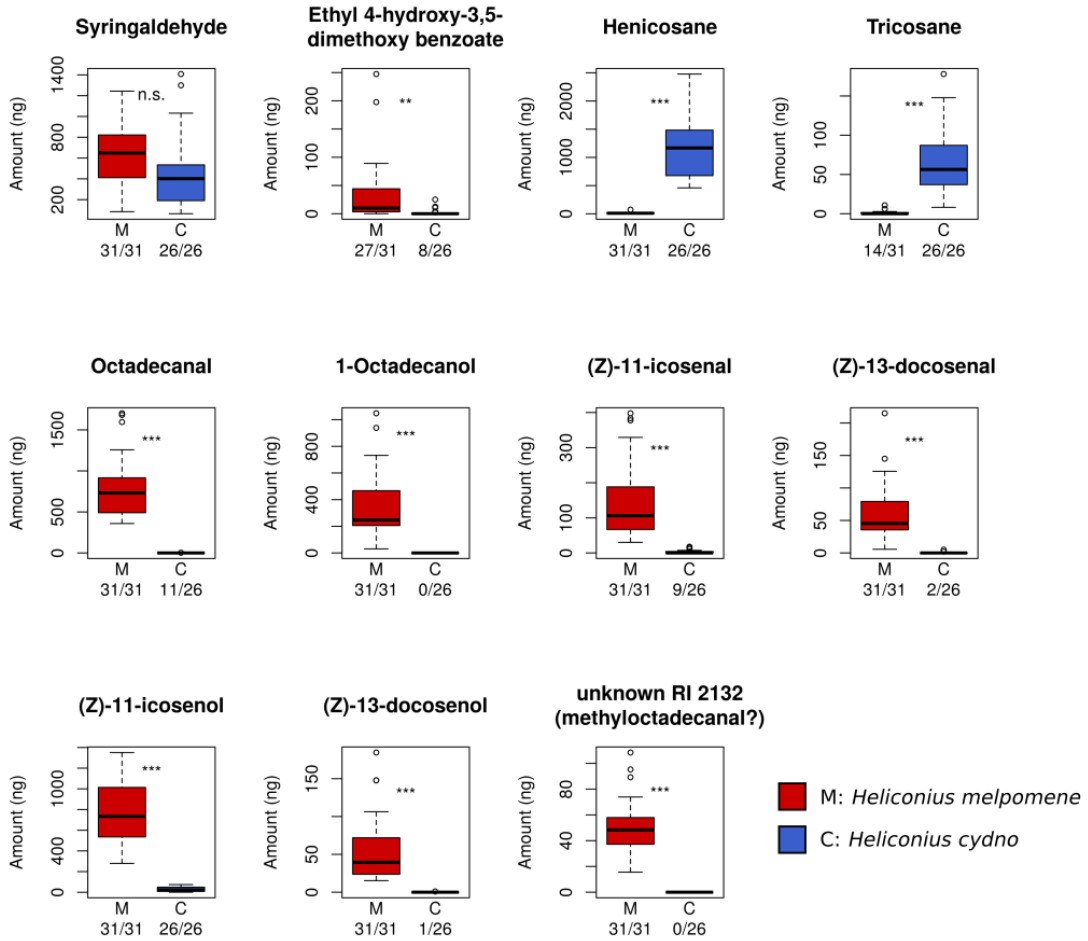
In general, LTA to *H. melpomene* pheromones was stronger than to *H. cydno* pheromones. Females of both species had different degrees of LTA to the two species' wing pheromones (in both cases, LTA was stronger to *H. melpomene* wings), while males had an equal degree of LTA. There was no obvious link between the presence or absence of LTA to a given pheromone component and the presence or absence of a given pheromone component in the preparation's species' wing pheromones. A significant correlation existed between degree of LTA and strength of response in to a given stimulus (Figure S8). Females of both species showed significant correlations for the synthetic stimulus sets and the overall data, while males of both species showed significant correlations between LTA strength and response strength for the natural stimulus sets and the overall data.

Short-term adaptation to pheromone extracts and synthetic compounds

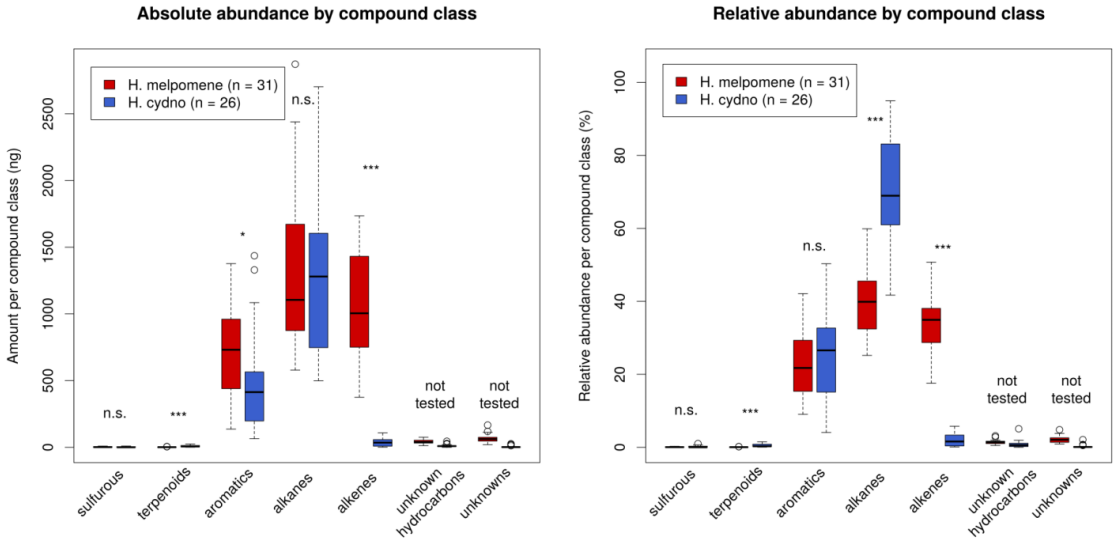
By contrast to LTA, short-term adaptation (STA) was mostly absent in all combinations of preparation species-sex and stimulus type (Table S4). No stimuli showed STA across all species-sex preparation combinations. Both female and male *H. melpomene* had STA to *H. cydno* wing pheromones. Female *H. cydno* also showed STA to *H. melpomene* wing extract. Female *H. melpomene* showed STA to 1-octadecanol, while male *H. cydno* showed STA to 1-octadecanol and (*Z*)-11-icosenal.

Supplemental Figures

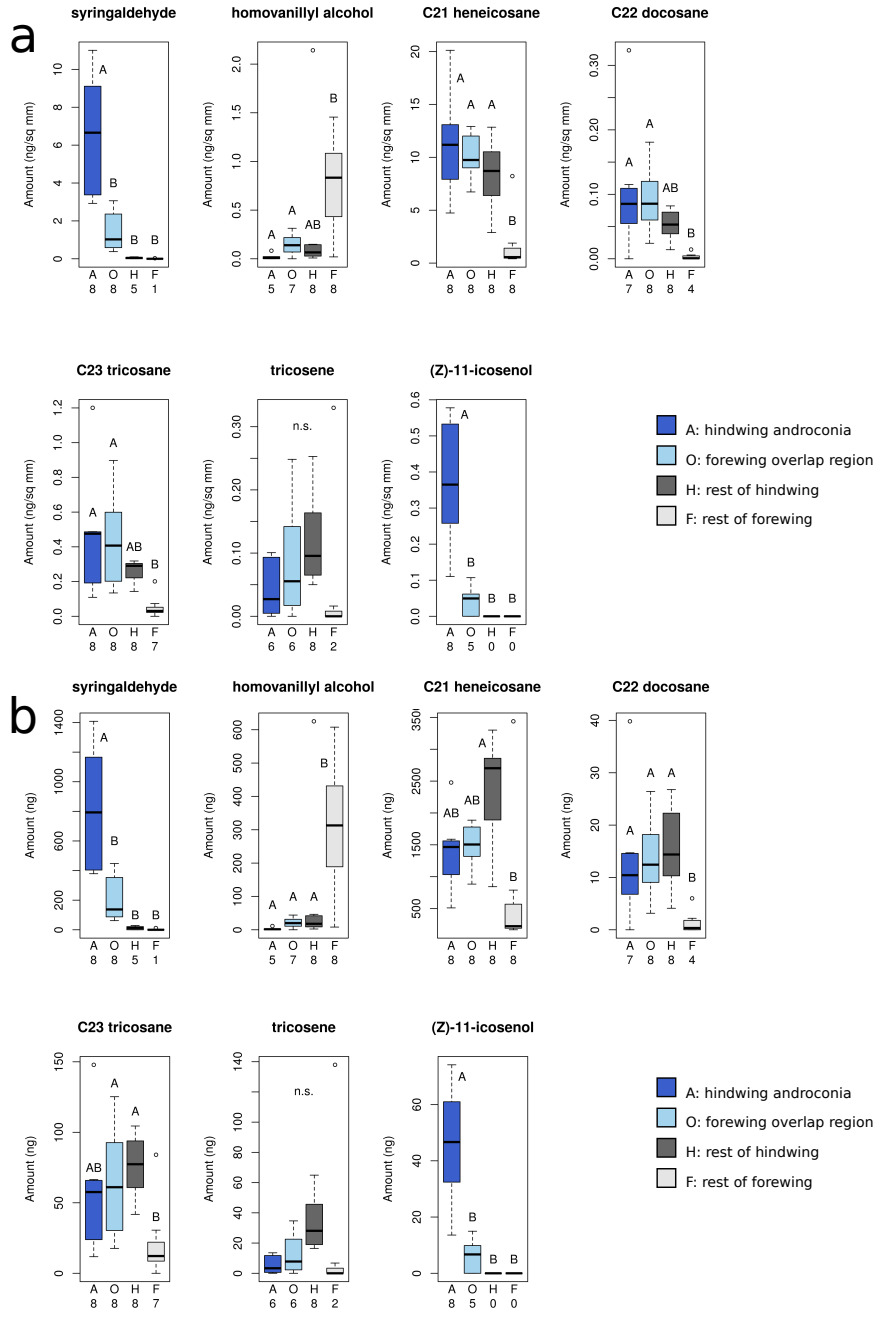
SI Figure 1: Comparison of the major compounds in *Heliconius melpomene* and *H. cydno*. Compounds shown are those contributing at least 1% of the total bouquet amount in either species. Numbers under each bar indicate how many samples (out of 31 for *H. melpomene* and 26 for *H. cydno*) the compound was found in. Significantly different compounds: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. n.s., not significant.



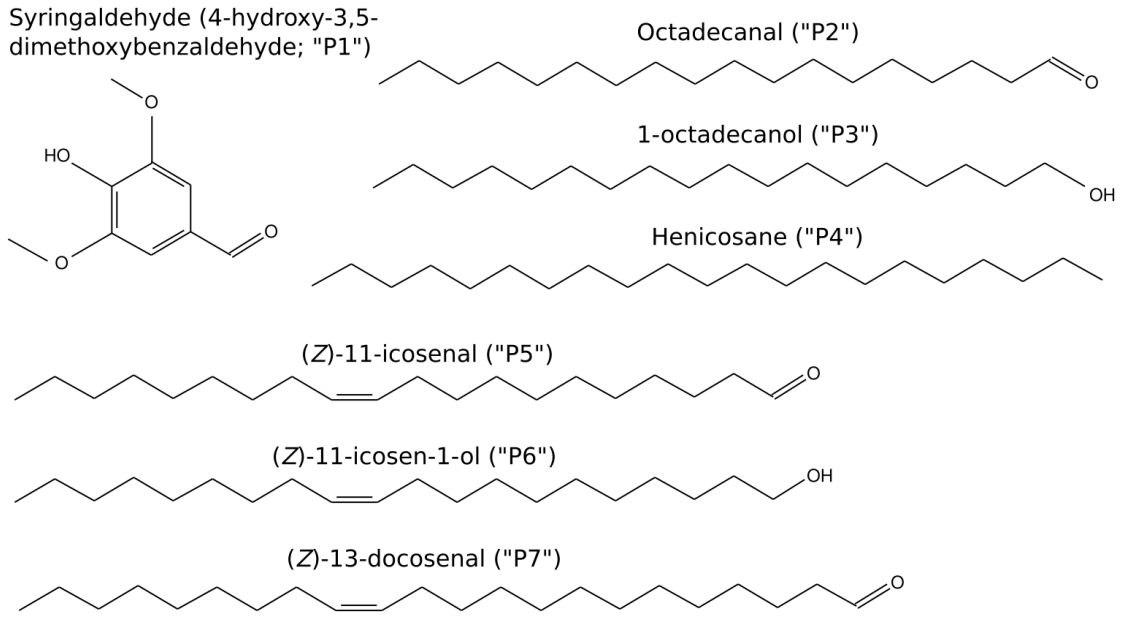
SI Figure 2: Absolute and relative abundance of different compound classes in *H. melpomene* and *H. cydno*. n.s., not significant; * p < 0.05; ** p < 0.01; *** p < 0.001. The two unknown categories were not tested as the compound types are not known.



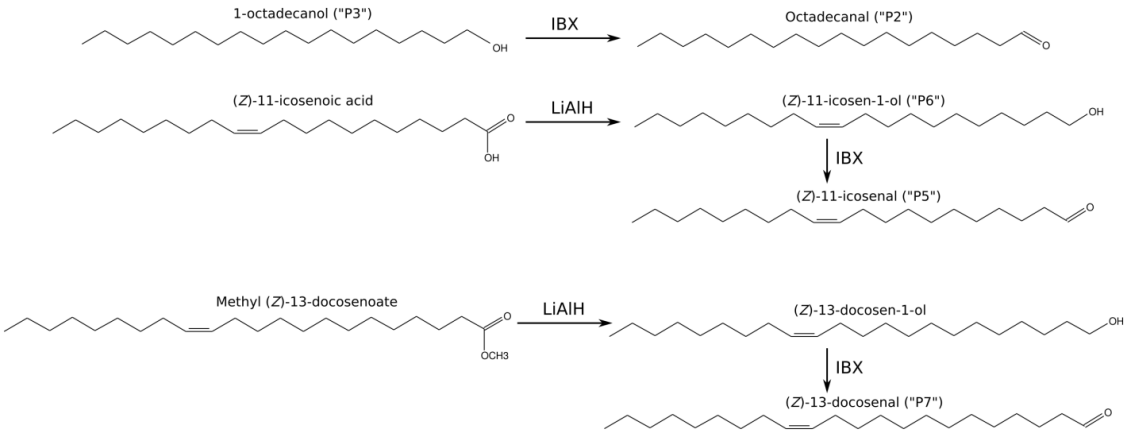
SI Figure 3: The seven compounds found in at least 0.1ng/mm^2 of wing tissue in at least one wing region in *Heliconius cydno*. a: Compound abundance per square millimeter of tissue. b: Compound abundance without tissue area correction. Numbers under each bar indicate how many samples (out of eight) the compound was found in; letters above bars indicate significant differences between regions. n.s., not significant. A: hindwing androconia; O: forewing overlap region; H: hindwing excluding androconia; F: forewing excluding overlap region.



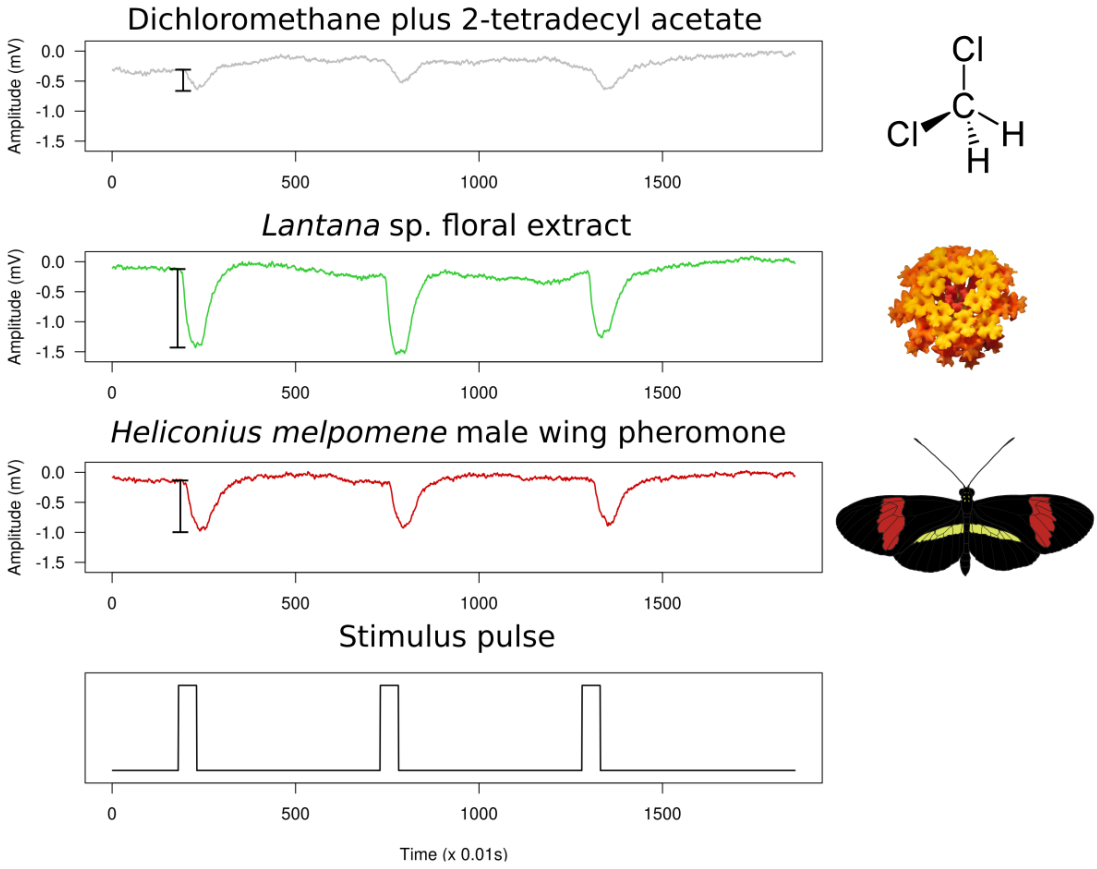
SI Figure 4: Structures and names of major components of the androconia of *H. melpomene* and *H. cydno* used in electrophysiological experiments.



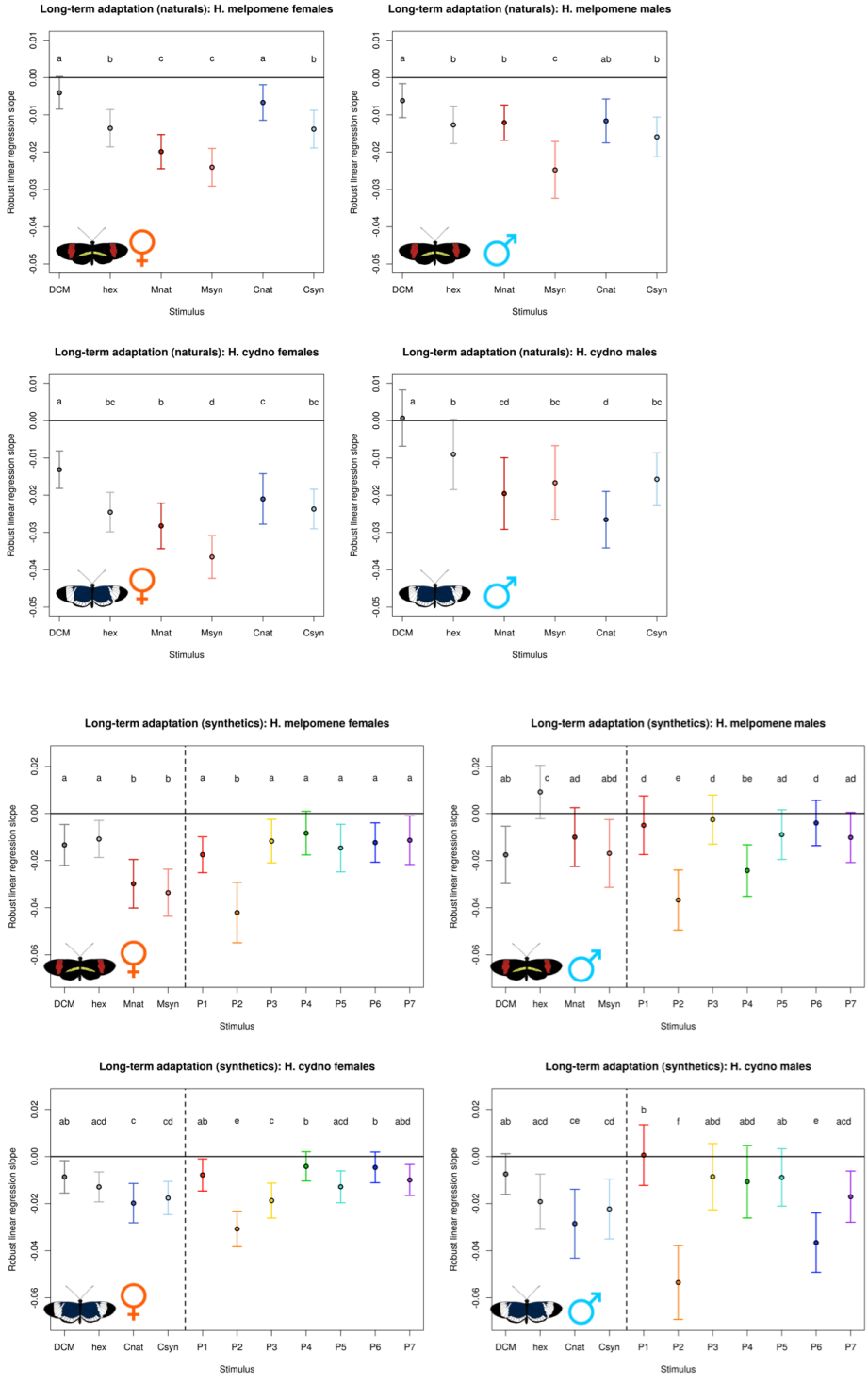
SI Figure 5: Synthesis of target compounds used in electrophysiological experiments. IBX: iodosobenzoic acid; LiAlH: lithium aluminum hydride.



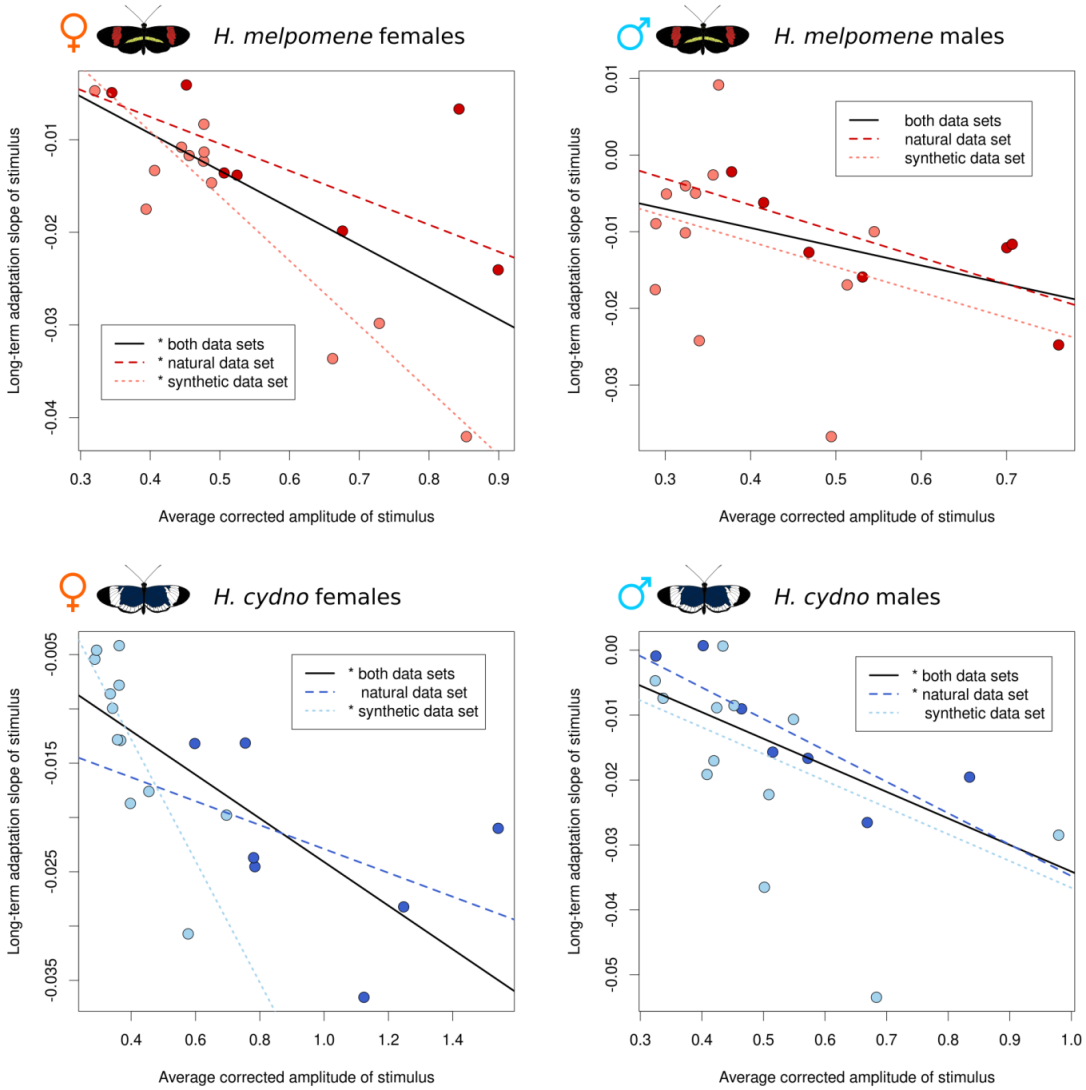
SI Figure 6: *Heliconius melpomene* responds to electrophysiological stimuli. Top to bottom: dichloromethane plus 2-tetradecyl acetate (internal standard) (negative control), *Lantana* extract (positive control), natural *H. melpomene* male wing extract, depiction of stimulus pulse timing. Data from a single virgin female. Bar-ended lines indicate the measured amplitude of the antennal response.



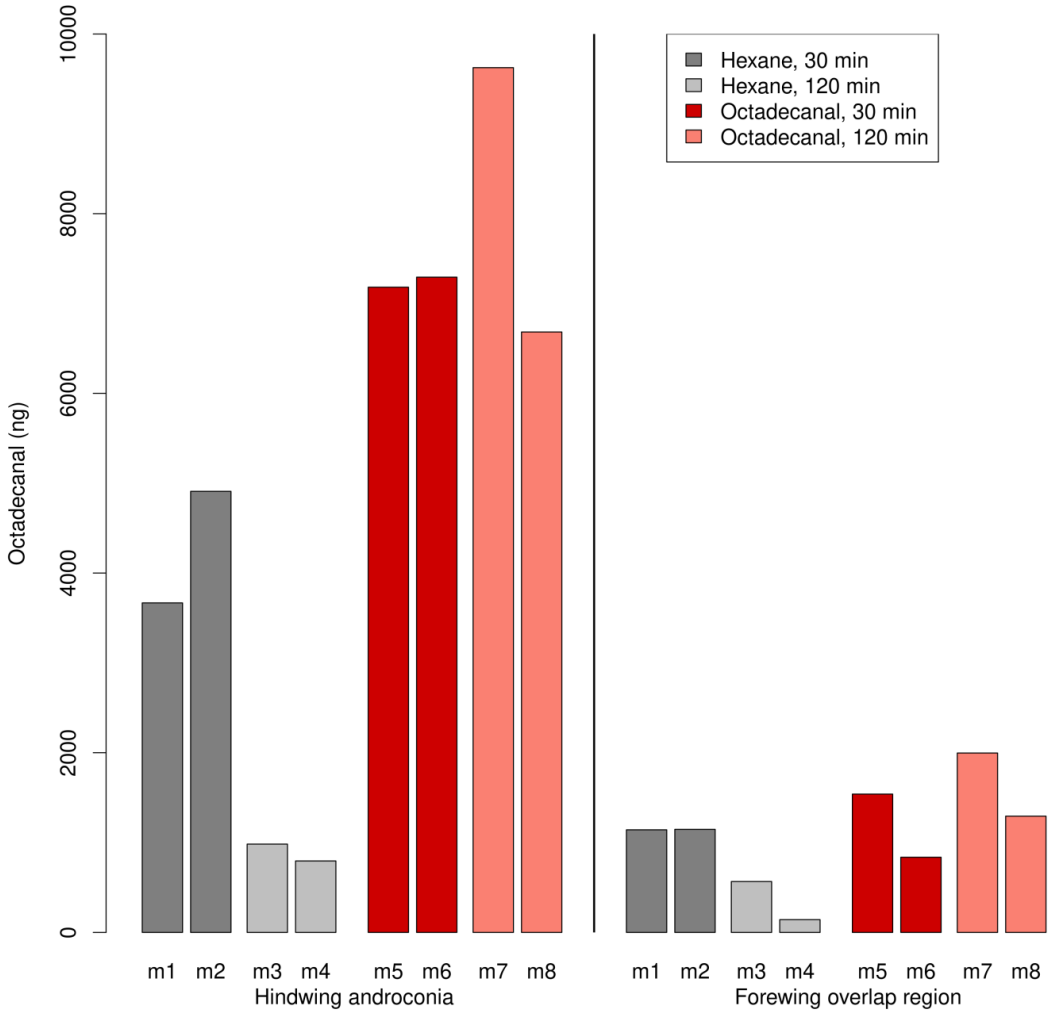
SI Figure 7: Long-term adaptation to natural and synthetic stimuli in *Heliconius* butterflies. The 95% confidence intervals of the robust LMM slope are shown; a negative slope means that responses to that stimulus drop over time. P1-P7: see Figure 2.



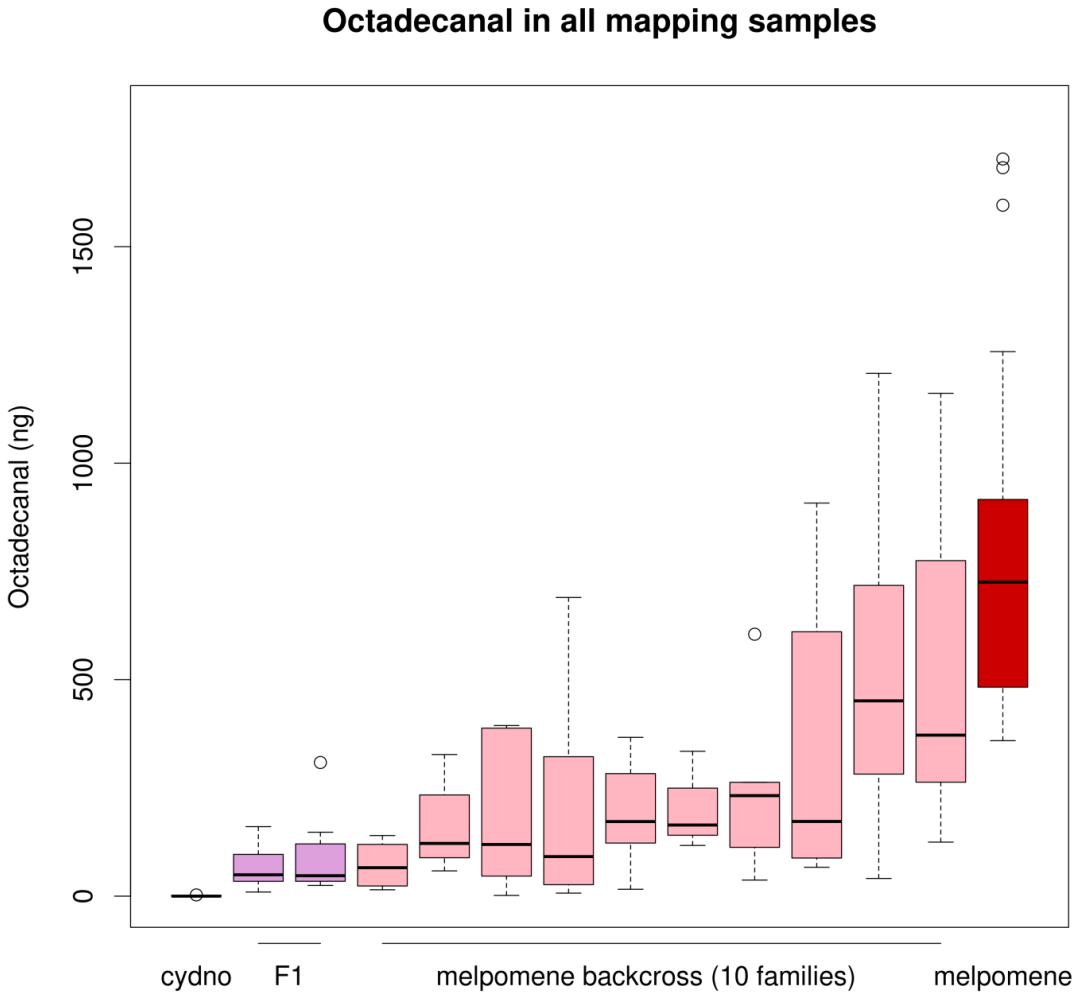
SI Figure 8: Strength of long-term adaptation correlates with amplitude of EAG response in a sex-specific fashion. In females, a stronger response to a given stimulus correlates with a stronger degree of LTA both overall and for the synthetic compound set. In males the same is seen overall and for the natural extract set in *H. cydno*, but not in *H. melpomene*.



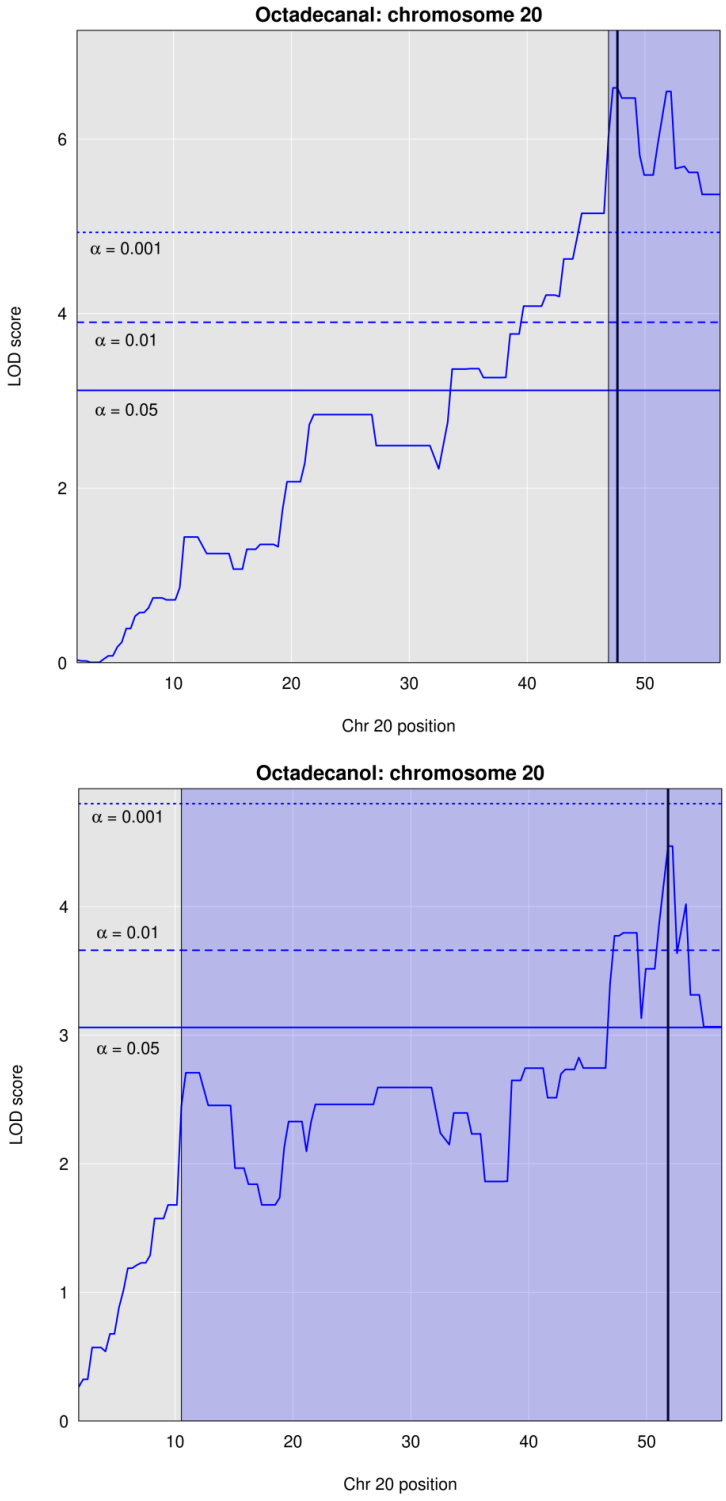
SI Figure 9: Octadecanal persistence in treated males over time. Bars show individual males, with two males per treatment-time point combination. Hindwing androconia and forewing overlap region males are shown in the same positions.



SI Figure 10: Octadecanal in *H. melpomene*, *H. cydno*, two F_1 families (one in each crossing direction), and the ten backcross to *H. melpomene* families used in QTL mapping. Colors: blue (*H. cydno*); purple (F_1 crosses of *H. melpomene* and *H. cydno*); pink (backcrosses to *H. melpomene*); red (*H. melpomene*).



SI Figure 11: Chromosome 20 QTL map for production of octadecanal and octadecanol in *H. melpomene*. Shaded regions indicate the Bayesian confidence intervals with kinship structure taken into account and black line indicates the peak of the QTL.



Supplementary Tables

SI Table 1: Comparison of *Heliconius cydno* and *H. melpomene* wing androconia bouquet. Amounts given are the mean in nanograms across 31 (*H. melpomene*) and 26 (*H. cydno*) samples, ± SD. Compounds are only included if found in at least 1/3 of samples for at least one species at levels of at least 1ng. #: compound found in at least 90% of samples of that species. Bold: species differ significantly. dof: Welch's two sample t-test degrees of freedom.

SI Table 2: Wing region specificity of *Heliconius cydno* androconial compounds. Amounts given are the mean across eight samples ± SD. Numbers in parentheses after compound amounts indicate the number of samples the compound was found in. Bold: wing regions differ significantly. NA: post-hoc test not conducted as original linear model not significant.

SI Table 3: Details of statistical comparisons of stimuli from electrophysiological experiments. P1-P7: as in the main text and Figure 4.

SI Table 4: Short-term adaptation to different stimuli in *H. melpomene* and *H. cydno* males and females. t1, t2, and t3: first, second, and third members of a stimulus triplet. dof: degrees of freedom. P1-P7: as in the main text and Figure 4.

SI Table 5: Potential candidate genes for octadecanal production underlying the QTL peak on chromosome 20 in *Heliconius melpomene*.

Supplementary tables available at (https://osf.io/saegq/?view_only=0634eba8ff8846c08ff7b87df769f04c)